

Fri Oct 12 16:09:17 2001

nmr nt

us-09-729-454-1.rsp

Seq. Alignment

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

QM protein - protein search, using SW model

Run on: October 12, 2001, 16:02:37 : Search time 16.9 Seconds

(without alignments)
962,802 Million cell updates/sec

Title: US-09-729-454-1
Percent score: 2485
Sequence: 1 MKISIMYKSLALLFLAS.....VHPQHYVGNQINILNYIC 475

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Search: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435
Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post processing: Minimum Match 08
Maximum Match 1008
Listing first 45 summaries

Database: SwissProt_39:*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Query Length	DB ID	Description
1	1606	64.6	540	1	BB61_RABIT
2	107.5	4.3	4563	1	APB_HUMAN
3	104	4.2	694	1	GLGH_HSVGU
4	103	4.1	793	1	SVL_MYCPN
5	101.5	4.1	1902	1	P2P_LACPA
6	101.5	4.1	3649	1	ACYS_MOCIA
7	100	4.0	1031	1	TEXT_EUPAE
8	98.5	4.0	802	1	CSO2_ECOLI
9	98	3.9	1174	1	NOS_RHOPI
10	97.5	3.9	271	1	NPI1_RAT
11	97.5	3.9	623	1	ABF2_HUMAN
12	97	3.9	988	1	PINH_ARATH
13	96.5	3.9	773	1	GLGH_SYNP7
14	96	3.8	770	1	RBP2_MOUSE
15	95.5	3.8	1102	1	RBP2_MOUSE
16	95	3.8	819	1	CSD1_ECOLI
17	95	3.8	1328	1	POLX_TORAC
18	93.5	3.8	454	1	GUAD_MOUSE
19	93.5	3.8	1132	1	DNBI_HSVGU
20	93.5	3.8	1616	1	VAED_SCHPO
21	93.5	3.8	1958	1	UBR1_SCHPO
22	92.5	3.7	1902	1	PIP_LACIC
23	92.5	3.7	538	1	VGIF_MOUSE
24	91	3.7	253	1	NPI1_MOUSE
25	91	3.7	698	1	SCNA_RAT
26	91	3.7	802	1	MOTS_BOCAI
27	91	3.7	1132	1	DNBI_HSVGU
28	90.5	3.6	627	1	CACP_CANTR
29	90.5	3.6	1577	1	MYSH_ACACA
30	90.5	3.6	1902	1	P2P_LACIA
31	90	3.6	538	1	VGIF_MOUSE
32	90	3.6	538	1	VGIF_MOUSE
33	90	3.6	610	1	MOTL_BORBU

34	89.5	3.6	569	1	Y139_MYCCE
35	89.5	3.6	668	1	P852_YEAST
36	89.5	3.6	839	1	NAH3_DIDMA
37	89.5	3.6	1108	1	CN3B_RAT
38	89.5	3.6	1857	1	FAS2_PENPA
39	89	3.6	382	1	VNAT_SVA1
40	89	3.6	538	1	VGIF_MOUSE
41	89	3.6	698	1	SCNA_MOUSE
42	89	3.6	908	1	DPOL_BORBU
43	89	3.6	989	1	VILI_DICDI
44	89	3.6	2787	1	TELI_YEAST
45	88.5	3.6	230	1	HBP_MAZE

ALIGNMENTS

RESULT	ID	BB61_RABIT	STANDARD	PRT	540 AA
AC	005004	BB61_RABIT	STANDARD	PRT	540 AA
DT	01-FEB-1994 (Rel. 28, Created)				
DT	01-FEB-1994 (Rel. 28, Last sequence update)				
DT	01-FEB-1996 (Rel. 33, Last annotation update)				
DE	BRUSH BORDER 61.9 KDA PROTEIN PRECURSOR.				
GN	ADRBAB-A.				
OS	Oryctolagus cuniculus (Rabbit).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OX	Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.				
RN	NCBI_TaxID:9986;				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=Intestine;				
RA	Medline:93286138; PubMed:8509424;				
RA	Boll W., Schmid-Chanda T., Semenza G., Mantel N.;				
RT	"Messenger RNAs expressed in intestine of adult but not baby rabbits.				
RT	Isolation of cDNA and characterization of a novel brush				
RT	border protein with esterase and phospholipase activity."				
RL	J. Biol. Chem. 268:12901-12911(1993).				
CC	-1- TISSUE SPECIFICITY: INTESTINE, AND IN A LESSER EXTENT IN KIDNEY				
CC	-1- DEVELOPMENTAL STAGE: EXPRESSED IN THE INTESTINE OF ADULT BUT NOT				
CC	BB61 RABBIT.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation				
CC	at the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/				
CC	or send an email to license@sib-sib.ch).				
CC	-----				
DR	EMBL: 212840; CAA78302.1; -				
DR	PIR: B45665; B45665.				
KW	Signal.				
FT	SIGNAL 1 22				
FT	CHAIN 23 540				
FT	SEQUENCE 540 AA; 61888 MW; EB65770EC69F25A CRC64;				
SO	POTENTIAL.				
SO	BRUSH BORDER 61.9 KDA PROTEIN.				
Query Match					
Best Local Similarity 64.6%; Score 1606; DB 1; Length 540;					
Matches 309; Conservative 67; Mismatches 91; Indels 72; Gaps 2;					
QY	8	YKSLALLFLASWIFVFNQSTKYSALNLSLHNNSTSLPPTPLSKPTE 67			
DB	3	HKYKLCILLAICVCLISQNSTKMGALPNHSYNSMISIPKSVSPKSLTE 62			
QY	68	TELKRELEKEDDOIPRPTHTVNTTATSTATILNPRDYTCRGDQLHLEVDHL 127			
DB	63	TELKRELEKEDDOIPRPTHTVNTTATSTATILNPRDYTCRGDQLHLEVDHL 122			
QY	128	GRRKQYGGDFLRNRSPPALMAGASGKVTDFNNGTYVSTFLWEOVSLSILHRSRG 187			

```

Db 123 GHOKEYGGDFLEAHMPSKAGASCKVTDFNNCTYLVSFTLFWEGQVSLVLLIHPSEG 182
Qy 188 VSALSNANQGYDRVFTGQFYNGTSQVHSEGLINTNAELCQVLDNRDQEGYCVRPQ 247
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 183 ASALARANQGYDRIIFKGQFYNGTSQVHSEGLINTNAELCQVLDNRDQEGYCVRPQ 242
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 248 HMPCAALFHYSKNKKVSVLSKQESKSPERSNGVSEIMKFKNTISVSKCN----- 297
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 243 HMPCEALFHYSTRNDSVILTSKEKLFHRSKVGVEIM-KNOHIDVSCQNKSEYKCKQ 301
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 298 ----- 297
Db 302 IGMKIPVPGVYTLQGRMLTTFNCTQIDTAKISGCKLGLIYLMGDSLTRQWIIYLPKYM 361
Qy 298 -TLKSVLDHESGKQLQHLAVIDRNRINQOKYCYPLIGSMYSVKEMEYLTRAIDRTGG 356
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
362 KTLKFEDLHETGNEFKKHLLEDAEKHTQLQMKKSHFVYQLFSVDHGVIPQIDRLIG 421
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 357 ENTVIVISLQHFRRFPIDVIRALNVHKAIOHLLRSPTWVIKTNIREMYNDAE 416
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 422 DKDVTIVITFQHFRRFPIDVIRRAISVQAERFLRSPTWKIVKTNIREMHIEAE 481
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 417 RFSDRHGYQVLIKIDFQDLVSVIDADMTIATGNNVHPQHVGVGNINILLNYIC 475
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 482 RFSDRHGYQVLIKIDFQDLVGVVDADMTIATGNNVHPQHVGVGNINILLNYIC 540
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 2
APB_HUMAN
ID APB_HUMAN STANDARD; PRT; 4563 AA.
AC P04114; O00502; Q13787;
DT 01-NOV-1986 (Rel. 03, Created)
DI 01-NOV-1986 (Rel. 03, Last sequence update)
DE 01-OCT-2000 (Rel. 40, Last annotation update)
DE APOLIPOPROTEIN B-100 PRECURSOR (APO B-100) [CONTAINS: APOLIPOPROTEIN
DE B-48 (APO B-48)].
GN APOB.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=870161385; PubMed=3763409;
   Knott T.C., Wallis S.C., Powell L.M., Pease R.J., Lusis A.J.,
   Blackhart B., McCarthy B.J., Mahley R.W., Levy-Wilson B., Scott J.;
   "Complete cDNA and derived protein sequence of human apolipoprotein
   B-100.";
RL Nucleic Acids Res. 14:7501-7503(1986).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=88003974; PubMed=3652907;
   Ludwig E.H., Blackhart B.D., Pierotti V.R., Calati L., Fortier C.,
   Knott T., Scott J., Mahley R.W., Levy-Wilson B., McCarthy B.J.;
   "DNA sequence of the human apolipoprotein B gene.";
RL DNA 6:363-372(1987).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=87008488; PubMed=3759943;
   Chen S.-H., Yang C.-Y., Chen P.-F., Setzer D., Tanimura M., Li W.-H.,
   Gotto A.M. Jr., Chan L.;
   "The complete cDNA and amino acid sequence of human apolipoprotein
   B-100.";
RL J. Biol. Chem. 261:12918-12921(1986).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=87041416; PubMed=3464946;
   Law S.W., Grant S.M., Higuchi K., Hospattankar A.V., Lackner K.J.,
   Lee N., Brewer H.B. Jr.;
   "Human liver apolipoprotein B-100 cDNA: complete nucleic acid and
   derived amino acid sequence.";
RL Proc. Natl. Acad. Sci. U.S.A. 83:8142-8146(1986).
RN [5]
RP SEQUENCE FROM N.A.
RX MEDLINE=87161758; PubMed=3030729;
   Cladaras C., Hadzopoulou-Cladaras M., Nolte R.T., Atkinson D.,
   Zannis V.I.;
   "The complete sequence and structural analysis of human
   apolipoprotein B-100: relationship between apoB-100 and apoB-48
   forms.";
RL EMBO J. 5:3495-3507(1986).
RN [6]
RP SEQUENCE OF 709-906 FROM N.A.
RX MEDLINE=85270450; PubMed=3860836;
   Deeb S.S., Motulsky A.G., Albers J.J.;
   "A partial cDNA clone for human apolipoprotein B.";
RL Proc. Natl. Acad. Sci. U.S.A. 82:4983-4986(1985).
RN [7]
RP SEQUENCE OF 3056-3159 FROM N.A.
RX MEDLINE=86041888; PubMed=3903660;
   Mehrabian M., Schumaker V.N., Fareed G.C., West K., Johnson D.F.,
   Kirchgesner T.G., Lin H.-C., Wang X., Ma Y., Mendiaz E., Lusis A.J.;
   "Human apolipoprotein B: identification of cDNA clones and
   characterization of mRNA.";
RL Nucleic Acids Res. 13:6937-6953(1985).
RN [8]
RP SEQUENCE OF 1937-2018 AND 3811-4334 FROM N.A.
RX MEDLINE=86093680; PubMed=3841204;
   Carlsson P., Olofsson S.O., Bondjers G., Darnfors C., Wiklund O.,
   Bjursell G.;
   "Molecular cloning of human apolipoprotein B cDNA.";
RL Nucleic Acids Res. 13:8813-8826(1985).
RN [9]
RP SEQUENCE OF 3109-4563 FROM N.A.
RX MEDLINE=85300528; PubMed=2994225;
   Knott T.J., Rall S.C. Jr., Innerarity T.L., Jacobson S.F.,
   Uredea M.S., Levy-Wilson B., Powell L.M., Pease R.J., Eddy R.,
   Nakai H., Byers M., Priestley L.M., Robertson E., Rall L.B.,
   Betsholtz C., Shows T.B., Mahley R.W., Scott J.;
   "Human apolipoprotein B: structure of carboxyl-terminal domains.
   sites of gene expression, and chromosomal localization.";
RL Science 230:37-43(1985).
RN [10]
RP SEQUENCE OF 1-291 FROM N.A.
RX MEDLINE=86149325; PubMed=3513177;
   Protter A.A., Hardman D.A., Schilling J.W., Miller J., Appleby V.,
   Chen G.C., Kirsher S.W., McEnroe G., Kane J.P.;
   "Isolation of a cDNA clone encoding the amino-terminal region of
   human apolipoprotein B.";
RL Proc. Natl. Acad. Sci. U.S.A. 83:1467-1471(1986).
RN [11]
RP SEQUENCE OF 1-1670 FROM N.A.
RX MEDLINE=86287319; PubMed=3461454;
   Protter A.A., Hardman D.A., Sato K.Y., Schilling J.W.,
   Yamanaka M., Hort Y.J., Hierrild K.A., Chen G.C., Kane J.P.;
   "Analysis of cDNA clones encoding the entire B-26 region of human
   apolipoprotein B.";
RL Proc. Natl. Acad. Sci. U.S.A. 83:5678-5682(1986).
RN [12]
RP PARTIAL SEQUENCE, AND IDENTIFICATION (APO-B48).
RX MEDLINE=88018019; PubMed=3659919;
   Chen S.-H., Habib G., Yang C.-H., Gu Z.-W., Lee B.R., Weng S.-H.,
   Silberman S., Cal S.-J., Deslypere J.P., Rosseneu M.,
   Gotto A.M. Jr., Li W.-H., Chan L.;
   "Apolipoprotein B-48 is the product of a messenger RNA with an organ-
   specific in-frame stop codon.";
RL Science 238:363-366(1987).
RN [13]
RP DOMAINS.
RX MEDLINE=87039351; PubMed=3773997;
   Knott T.C., Pease R.J., Powell L.M., Wallis S.C., Rall S.C. Jr.,
   Innerarity T.L., Blackhart B., Taylor W.R., Marcel V., Milne R.,
   Johnson D., Fuller M., Lusis A.J., McCarthy B.J., Mahley R.W.,
   Levy-Wilson B., Scott J.;
   "Complete protein sequence and identification of structural domains
   of human apolipoprotein B.";

```


Office

Applicant

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 22, 2002, 02:33:06 ; Search time 2189 Seconds
(without alignments)
11956.102 Million cell updates/sec

Title: US-09-729-454-3
Perfect score: 1616
Sequence: 1 atcattgctaaaggtaaca.....aatacaaaaaaaaaaaaaa 1616

Scoring table: OLIGO_MUC
Gapop 60.0 , Capext 60.0

Searched: 16154066 seqs, 8097743376 residues

Word size : 0

Total number of hits satisfying chosen parameters: 32308132

inimum DB seq length: 0
maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

EST:*

- 1: em_estba:*
- 2: em_esthum:*
- 3: em_estin:*
- 4: em_estmu:*
- 5: em_estov:*
- 6: em_estop:*
- 7: em_estro:*
- 8: em_hcc:*
- 9: gb_est1:*
- 10: gb_est2:*
- 11: gb_hcc:*
- 12: gb_est3:*
- 13: gb_est4:*
- 14: gb_est5:*
- 15: em_estfun:*
- 16: em_estom:*
- 17: gb_gss:*
- 18: em_gss_hum:*
- 19: em_gss_inv:*
- 20: em_gss_pin:*
- 21: em_gss_vrt:*
- 22: em_gss_fun:*
- 23: em_gss_mus:*
- 24: em_gss_mus:*
- 25: em_gss_other:*
- 26: em_gss_pro:*
- 27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
C 1	268	16.6	358	9	AI833131	AI833131 at75d06.x
C 2	169	10.5	905	12	BC199761	BC199761 RST19055
C 3	148	9.2	343	12	BC201797	BC201797 RST21139
C 4	112	6.9	112	9	AI581835	AI581835 ar94908.x
C 5	111	6.9	946	12	BC217603	BC217603 RST37315
C 6	109	6.7	530	17	AQ439960	AQ439960 HS_5059_B

7	88	5-4	634	13	BI764226
C 8	45	2-8	327	10	AW842215
9	32	2-0	564	10	BB625007
10	32	2-0	3267	11	AK019469
11	27	1-7	573	17	AZ454451
12	27	1-7	630	10	BB656245
13	24	1-5	378	12	BF721682
14	24	1-5	446	10	BB62365
15	24	1-5	597	12	BC088025
16	24	1-5	671	10	BB605016
17	23	1-4	208	14	BF75193
C 18	23	1-4	227	13	BG942595
C 19	23	1-4	252	10	AW522484
C 20	23	1-4	279	10	AW087776
C 21	23	1-4	357	9	AI136832
C 22	23	1-4	375	12	BC514625
C 23	23	1-4	403	9	AI482526
C 24	23	1-4	418	10	AW494429
C 25	23	1-4	418	13	BI884210
C 26	23	1-4	434	13	BI494352
C 27	23	1-4	436	9	AL047005
C 28	23	1-4	472	10	AW435156
C 29	23	1-4	490	13	BM445719
C 30	23	1-4	492	14	BQ018316
C 31	23	1-4	495	14	C93956
C 32	23	1-4	521	12	BG579237
C 33	23	1-4	536	12	BF510845
C 34	23	1-4	537	13	BM385280
C 35	23	1-4	544	14	BM989987
C 36	23	1-4	563	10	AW526128
C 37	23	1-4	564	13	BI402698
C 38	23	1-4	580	13	BM030140
C 39	23	1-4	588	14	BM973554
C 40	23	1-4	598	17	BH323298
C 41	23	1-4	620	14	BQ574501
C 42	23	1-4	622	9	A0039013
C 43	23	1-4	622	14	BM683810
C 44	23	1-4	623	14	BM971486
C 45	23	1-4	639	14	BQ190688

ALIGNMENTS

AI833131 358 bp mRNA linear EST 13-JUL-1999
at75d06.x1 Barstead colon HPLRB7 Homo sapiens CDNA clone
IMAGE:2377835.3, similar to SW:BB61_RABIT_Q05004 BRUSH BORDER 61.9
KD PROTEIN PRECURSOR. ;, mRNA sequence.

ACCESSION AI833131
VERSION AI833131
KEYWORDS EST.
SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 358)

AUTHORS

Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S., Krizman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin, J., Moore, B., Scheilenberg, K., Steptoe, M., Tan, F., Theising, B., White, F., Wylie, T., Waterston, R., and Wilson, R.

WashU-NCI human EST Project

Unpublished (1997)

CONTACT: Willson RK

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@wustl.wustl.edu

This clone is available royalty-free through LBNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.

Seq primer: -40UP from Gibco.

Fri Oct 12 16:09:17 2001

us-09-729-454-1.rsp

Page 1

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

1M protein - protein search, using sw model

Run on: October 12, 2001, 16:02:37 ; Search time 16.9 Seconds

(without alignments)
962.802 Million cell updates/sec

Title: US-09-729-454-1

Sequence: 2485
1 MKISMINKSLALLPLIAS.....VHPPOHYVGNQINILNYIC 475

Scoring table:

BLOSUM62
Gapop 10.0 , Capext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database: SwissProt_39.*

Fred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1606	64.6	540	1	BB61_RABIT
2	107.5	4.3	4563	1	APB_HUMAN
3	104	4.2	694	1	VGLE_HSV6U
4	103	4.1	793	1	STL_MYCPN
5	101.5	4.1	1902	1	P2P_LACPA
6	101.5	4.1	1649	1	ACVS_NOCIA
7	100	4.0	1031	1	TRPT_EUPAR
8	98.5	4.0	802	1	CSD2_ECOLI
9	94	3.9	1174	1	NBS_RHOPR
10	97.5	3.9	271	1	NPH1_RAT
11	97.5	3.9	623	1	ABF2_HUMAN
12	97	3.9	988	1	PIN1_ARATH
13	96.5	3.9	773	1	GLGB_SYNP7
14	96	3.9	770	1	RPOP_AGART
15	95.5	3.8	1102	1	CSD1_ECOLI
16	95	3.8	819	1	POIX_TOBAC
17	95	3.8	1328	1	GNAD_MOUSE
18	93.5	3.8	454	1	DNB1_HSV6U
19	94.5	3.8	1132	1	YABD_SCHPO
20	94.5	3.8	1616	1	UBR1_SCHPO
21	93.5	3.8	1958	1	P1P_LACIC
22	92.5	3.7	1902	1	VGLE_MOUSE
23	92	3.7	538	1	NPH1_MOUSE
24	91	3.7	253	1	SCAA_RAT
25	91	3.7	698	1	MOT5_BICAI
26	91	3.7	802	1	DNB1_HSV6U
27	91	3.7	1132	1	CACR_CANTR
28	90.5	3.6	627	1	MYSH_LACAC
29	90.5	3.6	1577	1	P2P_LACIA
30	90	3.6	1902	1	VGLE_MOUSE
31	90	3.6	538	1	VGLE_MOUSE
32	90	3.6	610	1	MUT1_BORBU
33	90	3.6	610	1	MUT1_BORBU

34	89.5	3.6	569	1	V139_MYCCE
35	89.5	3.6	668	1	PBS2_YEAST
36	89.5	3.6	839	1	NAH3_DIDMA
37	89.5	3.6	1108	1	CN3B_RAT
38	89.5	3.6	1857	1	FAS2_PENPA
39	89	3.6	382	1	VMAT_SVAL
40	89	3.6	538	1	VGLE_MOUSE
41	89	3.6	699	1	DPO1_MOUSE
42	89	3.6	908	1	DPO1_MOUSE
43	89	3.6	959	1	TELL_YEAST
44	89	3.6	2787	1	TELL_YEAST
45	89.5	3.6	230	1	HBP_MAZE

ALIGNMENTS

RESULT	1	BB61_RABIT	STANDARD	PRT	540 AA
AC	005004	BB61_RABIT			
DT	01-FEB-1994 (Rel. 28, Created)				
DT	01-FEB-1994 (Rel. 28, Last sequence update)				
DT	01-FEB-1996 (Rel. 33, Last annotation update)				
DE	BRUSH BORDER 61.9 KDA PROTEIN PRECURSOR.				
GN	ADRBAB-A.				
OS	Oryctolagus cuniculus (Rabbit).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.				
OX	NCBI_TaxID=9986;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=Intestine;				
RA	Boll W., Schmid-Chanda T., Semenza G., Mantel N.;				
RA	"Messenger RNAs expressed in Intestine of adult but not baby rabbits.				
RT	Isolation of cognate cDNAs and characterization of a novel brush				
RT	border protein with esterase and phospholipase activity."				
RL	J. Biol. Chem. 268:12901-12911(1993).				
CC	-1- TISSUE SPECIFICITY: INTESTINE; AND IN A LESSER EXTENT IN KIDNEY.				
CC	-1- DEVELOPMENTAL STAGE: EXPRESSED IN THE INTESTINE OF ADULT BUT NOT				
CC	BBAB RABBIT.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sdb.ch/announce/				
CC	or send an email to license@isb-sdb.ch).				
CC	-----				
DR	EMBL: 212840; CAY8302.1; ..				
DR	PIR: B45665; B45665.				
KW	SIGNAL.				
FT	CHAIN				
FT	SEQUENCE 540 AA: 61888 MW: 61888 MW: BB65770EC69P25A CRC64:				
SO	BRUSH BORDER 61.9 KDA PROTEIN.				
Query Match	64.6% Score 1606; DB 1; Length 540;				
Best Local Similarity	57.3% Pred. No. 6.3e-17;				
Matches 309; Conservative 67; Mismatches 91; Indels 72; Gaps 2;					
OY	8 YKSLALLPLIASLIIFFVFNQSKVMSALISLRYWNNSTSLFPTPLISKPLTE 67				
DB	3 HKYKLLCLLAACVLCIISONSRYKIGALKLPNSHYNSMSISSPKMSVPKSLTE 62				
OY	68 TELKLEIKLEKIDDOIPRPPTHTNTTSTSTTLINPDYCGGDLHLLEVRDHL 127				
DB	63 TELKLEIKLEKIDRLIPRPPTHTNTTSTSTTLINPDYCGGDLHLLEVRDYL 122				
OY	128 GRROVGGDFLARRSKPALMAGASGVTDPNNGYLVSEFLWEGVSLILLHPSEG 187				

Db 123 GHOKEGGDELFARAMPSPALAKAGSGKVTDPNNNGYLVSFLEMEQVSLVLLIHSEG 182
 Qy 188 VALMSAANGQYRYVFTGQCPNGTSQVHSECGLLTNALCOYLDNDQEGFCYCRPQ 247
 Db 183 ASALMRANNGQYRIIFKQCPNCTSHVFESCLTNSNTECKYLNGRDQVFCMKPQ 242
 Qy 248 HNPCCALTLHMYSNKKRVSTLSKQSKSLFERSNNGVEIENKENTISVSKCN----- 297
 Db 243 HNPCCALTLHVSINRDI SYLTSKEKNLFHRSKVGVEIN-KNOHIDVSGCNKSKVEKQ 301
 Qy 298 ----- 297
 Db 302 IOMKTEPVGTYLOGWMLTTCNOIQDPAKISGLCKKLLIYMGDSTLRQYTLPRVM 361
 Qy 298 -TLKSVLHESGKLOHQLAVDLDRNINIQMKYXCPYLGISMTYSVEMEYLTARITRG 356
 362 KTLKPFDEHETGEMFKHLLDLAEKHTQIOMKXHSRPFYTLQESVIDHGYIQEIDRLIG 421
 Qy 357 EKNTVYISLGGHFRPEPDIYFIRRALNHYKKAIOHLLRSDPTVNIKENTREMYNDA 416
 422 DMDYIVITFGHFRPEPDIYFIRRALSVQAEELFLRSATKIVKENTREHNEHAE 401
 Qy 417 RPSDEHGYIQLYIKIDFODLSVSTIDAMDITANGTNNVHPPOHYVGNQINILLNTIC 475
 Db 482 RFGDEHGYIQLYIKIDFODLNVGVVAMDNTIAGTNNVHPPOHYVGNQINILLNTIC 540
 RESULT 2
 APB_HUMAN STANDARD: PRT: 4563 AA.
 ID AC P04114: 000502: 013787:
 DT 01-NOV-1986 (Rel. 03 Created)
 DT 01-NOV-1986 (Rel. 03 Last sequence update)
 DT 01-OCT-2000 (Rel. 40 Last annotation update)
 DE APOLIPOPROTEIN B-100 PRECURSOR (Apo B-100) [CONTAINS: APOLIPOPROTEIN
 B-48 (Apo B-48)].
 GN APOB.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID:9606;
 RN 1)
 RP MEDLINE-87016385: PubMed-3763409:
 RZ Knott T.C., Wallis S.C., Powell L.M., Pease R.J., Lusis A.J.,
 Blackhart B., McCarthy B.J., Mahley R.W., Levy-Wilson B., Scott J.,
 Complete cDNA and derived protein sequence of human apolipoprotein
 B-100.*
 RL Nucleic Acids Res. 14:7501-7503(1986).
 RN 2)
 RP SEQUENCE FROM N.A.
 RZ MEDLINE-88003974: PubMed-3652907:
 RZ Ludwig E.H., Blackhart B.D., Plerotti V.R., Caiati L., Fortler C.,
 Knott T., Scott J., Mahley R.W., Levy-Wilson B., McCarthy B.J.,
 cDNA sequence of the human apolipoprotein B gene.*
 RL DNA 6:363-372(1987).
 RN 3)
 RP SEQUENCE FROM N.A.
 RZ MEDLINE-87008488: PubMed-3759943:
 RZ Chen S.-H., Yang C.-Y., Chen P.-F., Setzer D., Tanimura M., Li W.-H.,
 Goto A.M., Jr., Chan L.,
 *The complete cDNA and amino acid sequence of human apolipoprotein
 B-100.*
 RL J. Biol. Chem. 261:12918-12921(1986).
 RN 4)
 RP SEQUENCE FROM N.A.
 RZ MEDLINE-87041416: PubMed-3664946:
 RZ Law S.W., Grant S.M., Higuchi K., Hospattankar A.V., Lackner K.J.,
 Lee N., Brewer H.B., Jr.,
 *Human liver apolipoprotein B-100 cDNA: complete nucleic acid and
 derived amino acid sequence.*
 RL Proc. Natl. Acad. Sci. U.S.A. 83:8142-8146(1986).
 RN 5)
 RP SEQUENCE FROM N.A.
 RZ MEDLINE-87161758: PubMed-4030729:
 RZ Claerghs C., Hadzopoulou-Claerghs M., Nolte K., Aikens J.,
 Zannis V.I.,
 *The complete sequence and structural analysis of human
 apolipoprotein B-100: relationship between apoB-100 and apoB-48
 forms.*
 RL EMBO J. 5:3495-3507(1986).
 RN 6)
 RP SEQUENCE OF 709-906 FROM N.A.
 RZ MEDLINE-85270450: PubMed-3660836:
 RZ Deed S.S., Kotulsky A.G., Albers J.J.,
 A partial cDNA clone for human apolipoprotein B.
 RL Proc. Natl. Acad. Sci. U.S.A. 82:4983-4986(1985).
 RN 7)
 RP SEQUENCE OF 3056-3159 FROM N.A.
 RZ MEDLINE-86041888: PubMed-3903660:
 RZ Mehrbaban M., Schumaker V.N., Fareed G.C., West R., Johnson D.F.,
 Kirchgesner T.G., Lin H.-C., Wang X., Ma Y., Mendez E., Lusis A.J.,
 *Human apolipoprotein B: identification of cDNA clones and
 characterization of mRNA.*
 RL Nucleic Acids Res. 13:6937-6953(1985).
 RN 8)
 RP SEQUENCE OF 1937-2018 AND 3811-4334 FROM N.A.
 RZ MEDLINE-86093680: PubMed-3441204:
 RZ Carlsson P., Olofsson S.O., Bondjers G., Darnfors C., Wiklund U.,
 Bjursell G.,
 Molecular cloning of human apolipoprotein B cDNA.
 RL Nucleic Acids Res. 13:8813-8826(1985).
 RN 9)
 RP SEQUENCE OF 3109-4563 FROM N.A.
 RZ MEDLINE-85100528: PubMed-2294225:
 RZ Knott T.J., Rall S.C., Jr., Innerarity T.L., Jacobson S.F.,
 Udeas M.S., Levy-Wilson B., Powell L.M., Pease R.J., Eddy R.,
 Nakai H., Byers M., Priesley L.M., Robertson E., Rall L.B.,
 Rasmussen H., Shows T.B., Mahley R.W., Scott J.,
 *Human apolipoprotein B: structure of carboxyl-terminal domains,
 sites of gene expression, and chromosomal localization.*
 RL Science 230:37-43(1985).
 RN 10)
 RP SEQUENCE OF 1-291 FROM N.A.
 RZ MEDLINE-86149325: PubMed-3531377:
 RZ Procter A.A., Hardman D.A., Schilling J.W., Miller J., Appleby V.,
 Chen G.C., Kirshner S.W., McIntire G., Kane J.P.,
 *Isolation of a cDNA clone encoding the amino-terminal portion of
 human apolipoprotein B.*
 RL Proc. Natl. Acad. Sci. U.S.A. 83:1467-1471(1986).
 RN 11)
 RP SEQUENCE OF 1-1670 FROM N.A.
 RZ MEDLINE-86287319: PubMed-3461454:
 RZ Procter A.A., Hardman D.A., Sato K.Y., Schilling J.W.,
 Yamanaka M., Hort Y.J., Hierlieth K.A., Chen G.C., Kane J.P.,
 *Analysis of cDNA clones encoding the entire B-26 region of human
 apolipoprotein B.*
 RL Proc. Natl. Acad. Sci. U.S.A. 83:5678-5682(1986).
 RN 12)
 RP PARTIAL SEQUENCE, AND IDENTIFICATION (APO-B48).
 RZ MEDLINE-88018019: PubMed-3659919:
 RZ Chen S.-H., Habib G., Yang C.-H., Gu Z.-W., Lee B.P., Wang S.-H.,
 Silbermann S.R., Cal S.-Y., Desjardes J.P., Rosenau M.,
 Goto A.M., Jr., Li W.-H., Chan L.,
 *Apolipoprotein B-48 is the product of a messenger RNA with an organ
 specific in-frame stop codon.*
 RL Science 238:363-366(1987).
 RN 13)
 RP DOMAINS.
 RZ MEDLINE-87039351: PubMed-3773997:
 RZ Knott T.C., Pease R.J., Powell L.M., Wallis S.C., Rall S.C., Jr.,
 Innerarity T.L., Blackhart B., Taylor W.R., Marcel Y., Milne R.,
 Johnson D., Fuller M., Lusis A.J., McCarthy B.J., Mahley R.W.,
 Levy-Wilson B., Scott J.,
 *Complete protein sequence and identification of structural domains
 of human apolipoprotein B.*

Fri Oct 12 16:09:16 2001

us-09-729-454-1.ra1

Page 1

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

QM protein - protein search, using sw model

Run on: October 12, 2001, 16:02:37 ; Search time 20.85 Seconds

(without alignments)
469,085 Million cell updates/sec

Title: US-09-729-454-1

Perfect score: 2485

Sequence: 1 MKISMINKSLILFLILAS.....VHPQHVGNQINILNLATIC 475

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 197339 seqs, 20590346 residues

Total number of hits satisfying chosen parameters: 197339

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Data base :

Issued Patents: AA: *
1: /cgn2_6/ptodata/2/1aa/5a.COMB.pep: *
2: /cgn2_6/ptodata/2/1aa/5a.COMB.pep: *
3: /cgn2_6/ptodata/2/1aa/6a.COMB.pep: *
4: /cgn2_6/ptodata/2/1aa/6a.COMB.pep: *
5: /cgn2_6/ptodata/2/1aa/6a.COMB.pep: *
6: /cgn2_6/ptodata/2/1aa/6a.COMB.pep: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	104	4.2	1494	3	US-08-755-587-186
2	100	4.0	1007	3	US-08-851-843A-86
3	100	4.0	1007	4	US-08-974-549A-187
4	100	4.0	1007	4	US-08-851-843A-2
5	100	4.0	1031	4	US-08-974-549A-110
6	95	3.8	610	2	US-08-879-561-10
7	94.5	3.8	1579	3	US-08-755-587-184
8	92	4.7	849	3	US-08-804-439A-17
9	92	3.7	849	3	US-08-720-229-17
10	90.5	3.6	2325	4	US-08-417-089-6
11	90.5	3.6	2325	4	US-08-695-651-6
12	90.5	3.6	2325	4	US-08-930-285-6
13	90	3.6	698	1	US-08-376-362A-20
14	89.5	3.6	668	1	US-08-530-950-13
15	89.5	3.6	668	4	US-09-149-879-13
16	87.5	3.5	315	1	US-08-571-758-12
17	87.5	3.5	315	1	US-08-909-984A-12
18	87.5	3.5	315	1	US-08-909-983-12
19	87.5	3.5	648	1	US-08-185-282-4
20	87.5	3.5	1418	2	US-08-639-501-2
21	87.5	3.5	3418	3	US-09-044-946-2
22	87.5	3.5	3418	3	US-09-044-908-2
23	86.5	3.5	648	1	US-08-185-282-3
24	86.5	3.5	2329	3	US-08-755-587-16
25	86	4.5	395	1	US-07-931-943-5
26	85.5	3.4	346	1	US-08-276-151-5
27	85.5	3.4	648	1	US-08-276-151-2

28	85.5	3.4	648	1	US-08-185-282-12	Sequence 12, Appl
29	85.5	3.4	648	2	US-08-886-751A-6	Sequence 6, Appl
30	85.5	3.4	648	3	US-09-209-668-13	Sequence 13, Appl
31	85.5	3.4	3418	2	US-08-603-753D-4	Sequence 4, Appl
32	85.5	3.4	3418	3	US-08-755-587-44	Sequence 44, Appl
33	85.5	3.4	3418	4	US-09-099-753-4	Sequence 4, Appl
34	85.5	3.4	3418	4	US-08-986-106-4	Sequence 4, Appl
35	84.5	3.4	387	4	US-09-323-427-4	Sequence 4, Appl
36	84.5	3.4	521	2	US-08-737-825-3	Sequence 3, Appl
37	84.5	3.4	648	1	US-08-185-282-1	Sequence 1, Appl
38	84.5	3.4	648	1	US-08-185-282-2	Sequence 2, Appl
39	84	3.4	417	1	US-08-351-981-7	Sequence 7, Appl
40	83	3.3	395	2	US-07-931-943-2	Sequence 2, Appl
41	83	3.3	395	2	US-08-624-601-15	Sequence 15, Appl
42	83	3.3	417	1	US-08-351-981-6	Sequence 6, Appl
43	83	3.3	874	2	US-08-677-734A-11	Sequence 11, Appl
44	83	3.3	874	3	US-08-804-439A-15	Sequence 15, Appl
45	83	3.3	874	3	US-08-720-229-15	Sequence 15, Appl

ALIGNMENTS

RESULT 1
US-08-755-587-186
Sequence 186, Application US/08755587
Patent No. 6045997
GENERAL INFORMATION:
APPLICANT: Futrel, Phillip A
APPLICANT: Wooster, Richard F
APPLICANT: Ashworth, Alan
APPLICANT: Stratton, Michael R
TITLE OF INVENTION: Materials and methods relating to the
TITLE OF INVENTION: Identification and sequencing of the BRCA2 cancer
NUMBER OF SEQUENCES: 222
CORRESPONDENCE ADDRESS:
ADDRESSEE: Bell Seitzer Park & Gibson
STREET: 310 UCB Plaza, 3605 Glenwood Avenue, PO Drawer 31107
CITY: Raleigh
STATE: NC
COUNTRY: USA
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/755,587
FILING DATE: 25-NOV-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/755,587
FILING DATE: 23-NOV-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 952555.0
FILING DATE: 14-DEC-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9617961.9
FILING DATE: 28-AUG-1996
ATTORNEY/AGENT INFORMATION:
NAME: Kenneth D Sibley
REGISTRATION NUMBER: 31,665
REFERENCE/DOCKET NUMBER: 5405-135
INFORMATION FOR SEQ ID NO: 186:
SEQUENCE CHARACTERISTICS:
LENGTH: 1494 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-755-587-186
Query Match 4.2%, Score 104; DB 3; Length 1494;
Best Local Similarity 21.3%; Pred. No. 0.3;

Matches 79; Conservative 54; Mismatches 142; Indels 96; Gaps 19;

```

QY 22 IIFVFNSTKYWSALNLSLHYMNNSTKSLFPTKPLSLK-----PLTTEL- 70
DB 432 IISGFQNSTY-----SDSESGHAPPTLSLKQDDSRNLTFSQAKET 477
QY 71 RIKEIIEKIDQIPRPFP---THV---NTTSATHTATILNPRDYICRGDOLHILLEVD 125
DB 478 ELSTLIEESGSEFFETQFRKPSHIQKNPFMPENQTLINSTKEMMDLHITTMAS 537
QY 126 -HLGRKQ---YGGDFLRARMSPALMAGASGVTDENNQTYLVEFTLFMCQ-VSLSL 180
DB 538 IQYDSKSKSGIIGKQKFACLSTSONSASGYSTDKN---VEFGFYSARQTKLV- 592
QY 181 LHPSEGVSAIWSAR-----NGYDRVIFGTQGFVNGTSQY----- 215
DB 593 -----GSEALQKAKKLFSDLNINDETSEVDRSFSSKYNDVSMTIQIDCNKNPE 646
QY 216 HSEGLINTNAELC-----QYLDNRDQGFYCVRPQHMCALT--HMYSKNNK 263
DB 647 NNKCRLLIONNIEMTIDIPVEETESYKRNTEENGNQCDADRNTCSSESQSSKNDT 706
QY 264 VSYLSKQKSL--PERSNGVEIMKRF---NT-----ISVSKNTLKSVDL-HESGKLQ 311
DB 707 V-YIHEENGLPICIDQHNIDKLFSQFMKEGNTQIKESLDTLCLVMKAETSHVTMSK 765
QY 312 HOLAVDLDRNI 322
DB 766 QOLTANTGONI 776

```

RESULT 2

```

US-08-851-843A-86
; Sequence 86, Application US/08851843A
; Patent No. 6093809
; GENERAL INFORMATION:
; APPLICANT: Cech, Thomas R.
; APPLICANT: Lingner, Joachim
; APPLICANT: Nakamura, Toru
; APPLICANT: Chapman, Karen B.
; APPLICANT: Morin, Gregg B.
; APPLICANT: Hatley, Calvin
; APPLICANT: Andrews, William H.
; TITLE OF INVENTION: No. 6093809el Telomerase
; NUMBER OF SEQUENCES: 225
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, 8th floor
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/851,843A
; FILING DATE: 06-MAY-1997
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/846,017
; FILING DATE: 25-APR-1997
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/844,419
; FILING DATE: 18-APR-1997
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/724,643
; FILING DATE: 01-OCT-1996
; CLASSIFICATION:

```

```

; ATTORNEY/AGENT INFORMATION:
; NAME: Apple, Randolph T.
; REGISTRATION NUMBER: 36,429
; REFERENCE/DOCKET NUMBER: 015369-0029300S
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0300
; TELEFAX: (415) 576-0200
; INFORMATION FOR SEQ ID NO: 86:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1007 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-851-843A-86

```

Query Match 4.0%; Score 100; DB 3; Length 1007;
Best Local Similarity 21.5%; Pred. No. 0.4;

Matches 106; Conservative 71; Mismatches 179; Indels 137; Gaps -

```

QY 18 LASWITFVFNSTKYWSALNLSLHYMNNSTKSLFPTKPLSLKILHTELK---K 79
DB 29 LYSWI-----QKVICRNSQS-HKLEDEIKIFAQTNIVAPRDYNEELFKVIANK 79
QY 74 EL-----IEKIDQIIPRPFTVNTTSATHTATILNPRDYICRGDOL--HILLEVA 124
DB 80 EYFSTGIMTELIDKCL-----VELLSGSVSDRQLVGFQFQKQVLAETHLLTA-- 140
QY 125 DHLGRKQYGDFLRARMSPALMAGASGVTDENNQTYLVEFTLFMCQVSLIHLH 184
DB 131 --LSTQKQY--PFQDENQVAMIG-----NELFRHLYTKY-----LIFORT 184
QY 185 SEGVSAIWSARNGYDRVIFGTQF---VNGTSQVH---SEGLINTNAEL-VYLFN 244
DB 169 SEGTVQPCGNV-FDHLKVNDRKQKGAADNMPKQSTCK--YNNKNDKDHFLNN 245
QY 236 RDQEG-----FYCVRPQHMCALTTHYSKNNKQSYLSQKSLPERSNGVEIM 285
DB 226 IIVFNWNNMKSRTIRYC-----THF---NRNNQPFKKEH--FVSKNNISAM 286
QY 286 EKPNFI--SVSKNTL-KSYDLHESGKLOHVLAVDLDRKINIQOKVYTLISMTYSVK 344
DB 265 DRAQITTNIFRNRIKRLKLDKVIETIAVLEKVDENFNYLTQSG-PLAENMEREKQ 327
QY 343 EMEVLTRAIDRTGK-----NTVVISLQGHF-RPPIDVPIR----- 381
DB 328 KIENL---INKTREESKYBELFSYTDKCVTQFINPFYNIILPDLTGNNRNKQK 384
QY 382 -----ALNYHKAIOHLLSRPDTWYI---IKTENIREMYNDAERFSDFHGYIYLI 430
DB 385 KVKKYVELNKHLEIKHNLLEKTIETREISMVQETSAKHYTFDHEH-----VYLMKLL 440
QY 431 KDIFQDLSVSI 442
DB 440 RWPFEDLVSLI 451

```

RESULT 3

```

US-08-974-549A-187
; Sequence 187, Application US/08974549A
; Patent No. 6166178
; GENERAL INFORMATION:
; APPLICANT: Cech, Thomas R.
; APPLICANT: Lingner, Joachim
; APPLICANT: Nakamura, Toru
; APPLICANT: Chapman, Karen B.
; APPLICANT: Morin, Gregg B.
; APPLICANT: Hatley, Calvin
; APPLICANT: Andrews, William H.
; TITLE OF INVENTION: Human telomerase catalytic subunit
; NUMBER OF SEQUENCES: 727
; CORRESPONDENCE ADDRESS:

```


Fri Oct 12 16:09:16 2001

us-09-729-454-1.ral

Page 3

```

ADDRESS: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/974,549A
FILING DATE: 19-NOV-1997
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/724,643
FILING DATE: 01-OCT-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/844,419
FILING DATE: 18-APR-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/846,017
FILING DATE: 25-APR-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/851,843
FILING DATE: 06-MAY-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/854,050
FILING DATE: 09-MAY-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/911,312
FILING DATE: 14-AUG-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/912,951
FILING DATE: 14-AUG-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/915,503
FILING DATE: 14-AUG-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO FCT/US97/17618
FILING DATE: 01-OCT-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO FCT/US97/17885
FILING DATE: 01-OCT-1997
ATTORNEY/AGENT INFORMATION:
NAME: Apple, Handolph Ted
REGISTRATION NUMBER: 36,429
REFERENCE/DOCKET NUMBER: 015389-00261005
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 187:
SEQUENCE CHARACTERISTICS:
LENGTH: 1007 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-974-549A-187

Query Match 4.0%; Score 100; DB 4; Length 1007;
Best Local Similarity 21.5%; Pred. No. 0.4;
Matches 106; Conservative 71; Mismatches 179; Indels 136; Gaps 29;
UY 16 LASWITFVONSTKYWSALNLSLHWNNSTKSLPKPTPLISL-KPLETETELRI---K 73
DB 29 LYSWI-----QKVICRNOSS-HYKDEDIKIPAOQTNIVATPRDYNEDEPKVIARK 79
QY 74 RI-----TEKLDQAIPIPPFTHVNTTTSATSHSTATILNPRDTYCGRGQUL---HILLEVR 124
DB 80 EVESTGLMIFELIDKCL-----VELLSSTIVSDRKLQCFGLKGNQAKTHLTLTA-- 130
```

```

QY 125 DHGRRKXGSDFLRARMSSPALMACASQKVTDFNNGTYLVSTFLFMEGQVSLSLIHP 184
DB 131 --LSTOKOY---FQDEMNQVARMIG-----NELFHHLYTY-----LIFORT 168
QY 185 SEGVSAIWSARNQGYDVIYFTGF---VNGTSQVH-----SECGLIINTNAELQOYLDN 235
DB 169 SEGVLVQFCGNVY-FDHLKYNDFDKQKGAADNPNRCSCTCK--YNNKNDHFLNN 225
QY 236 RDEEG-----FYCVRPQHMPALTHWYSKNKKVSYLSKQEKSLFERSVGYEIN 285
DB 226 INPNNNNNKKSRIRFLYC-----THP---NNNNOFRKHE---FVSNNNNSAM 268
QY 286 EKENTL--SVSKCNTL-KSYDLHESGKLQHQLAVDLDRNINIQWOKYCPPLIGSMYSVK 342
DB 269 DRAQTEFTNIFEFNRIRKRLKDKVIEKIAVMLEKVFDFNNYTLKSC-PLPENWREKQ 327
QY 343 EMEYLFRAIDRTGEEK-----NTVIYISLQHF-RPPIDVIFIR----- 381
DB 328 KIENL---INKTREKSKYEEELFSTTDNKCVCVQFINFEFTNILDKDFLGRNRKFK 384
QY 382 -----ALNVKAI-QHLLLRSPDTNVI---IKTENIREMYNDAERFSDPHGYOYLIT 430
DB 385 KVKKYVELNKHHLHKMLLEKINTREISMVQVETSAKHFFYFDHENDI-----YVLWKL 439
QY 431 KDIFFODLSVSI 442
DB 440 RMIFEDLVYSLI 451

RESULT 4
US-08-851-843A-2
Sequence 2, Application US/08851843A
Patent No. 6093809
GENERAL INFORMATION:
APPLICANT: Cecch, Thomas R.
APPLICANT: Lingner, Joachim
APPLICANT: Nakamura, Toru
APPLICANT: Chapman, Karen B.
APPLICANT: Morin, Gregg B.
APPLICANT: Harley, Calvin
APPLICANT: Andrews, William H.
TITLE OF INVENTION: No. 6093809e1 Telomerase
NUMBER OF SEQUENCES: 225
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, 8th Floor
CITY: San Francisco
STATE: California
COUNTRY: United States of America
ZIP: 94111
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/851,843A
FILING DATE: 06-MAY-1997
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/846,017
FILING DATE: 25-APR-1997
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/844,419
FILING DATE: 18-APR-1997
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/724,643
FILING DATE: 01-OCT-1996
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
```

NAME: Apple, Randolph T.
 REGISTRATION NUMBER: 36,429
 REFERENCE/DOCKET NUMBER: 015389-002930US
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (415) 576-0200
 TELEFAX: (415) 576-0300
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1031 amino acids
 TYPE: amino acid
 STRANDEDNESS: not relevant
 TOPOLOGY: not relevant
 MOLECULE TYPE: protein
 US-08-851-843A-2

Query Match 4.0%; Score 100; DB 3; Length 1031;
 Best Local Similarity 21.5%; Pred. No. 0.41;
 Matches 106; Conservative 71; Mismatches 179; Indels 136; Gaps 29;

QY 18 LAMWIFTFQNSTKVMASLNLSLHYNNSTKSLFKPTPLSL-KPLETELRT---K 73
 Db 30 LYSWI-----QKVICRNSQS-HIKDLEDIKIFAQTNIVATPRDYNEEDPKVIARK 80
 QY 74 EI-----IEKLDQIIPPRFTHTVTTTSATHTATILNPRDYCRGDQ---HILLEVR 124
 Db 81 EVFSTGLMIELIDKCL-----VELLSSDVSVDROKLCFCGFCQKNGQLAKTHLLTA-- 131
 QY 125 DHLGRKQYGGDFLRARMSPALMAGASGVTFDFNNGTYLVSFTLFWEQVSLLLHP 184
 Db 132 --LSTQKQY--FFQDEMNQVAMIG-----NELFRHLVTKY-----LIFORT 169
 QY 185 SEGVSAALMSARNOGYRVYFTQGF---VNGTSQVH---SEGLILNTNMLCOYLDN 235
 Db 170 SEGTLVQFCGNV-FDHLVNDKFDKQKGGADNNEPCSTCK--YNNKKNKHQFLNN 226
 QY 236 KDQES-----FYCVRPQHPCALTHMYSKKKKYSYLSEKSLPERSNNGVEIM 285
 Db 227 INVPMNMNKSRTIRFYC-----THF---NRNNGFFKHE---FVSKNNNISAM 269
 QY 286 EKFNFI--SVSKCNTL-KSYDLHESGKLOHQLAVDLDRNINIQOKCYPLIGSYTVSK 342
 Db 270 DDAQIIFITNIFRNRIKKLKDVKYIKIATMEKKYKDFNFYVLTQSC-PLDENMRERKQ 328
 QY 343 EMEYLTALDRTGGEK-----NTVVISLQGH-REFPLDYFIR----- 381
 Db 329 KIENL---INKTREKSKYEELEFYTDMKCVTQFINFEFYNILPRDELTRNNKKNQK 385
 QY 382 -----ALNFKAI-OHLLSRPDTMYI---IKTENREKYNDAERFSDPHGYQYI 430
 Db 386 KYKKYVELNKHLEIKNLLKINTREISMQVETSAKHFFYEDHENI-----VYLKRL 440
 QY 431 KDFQDLVSII 442
 Db 441 KPIFEDLWVSLI 452

RESULT 5
 US-08-974-549A-110
 ; Sequence 110, Application US/08974549A
 ; Patent No. 6166178

GENERAL INFORMATION:
 APPLICANT: Cech, Thomas R.
 APPLICANT: Lingner, Joachim
 APPLICANT: Nakamura, Toru
 APPLICANT: Chapman, Karen B.
 APPLICANT: Morlin, Gregg B.
 APPLICANT: Hartley, Calvin B.
 APPLICANT: Andrews, William H.
 TITLE OF INVENTION: Human telomerase catalytic subunit
 NUMBER OF SEQUENCES: 727
 CORRESPONDENCE ADDRESSES:
 ADDRESSEE: Townsend and Townsend and Crew LLP

STREET: Two Embarcadero Center, Eighth Floor
 CITY: San Francisco
 STATE: California
 COUNTRY: USA
 ZIP: 94111-3834
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/974,549A
 FILING DATE: 19-NOV-1997
 CLASSIFICATION: 536
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/724,643
 FILING DATE: 01-OCT-1996
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/844,419
 FILING DATE: 18-APR-1997
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/846,017
 FILING DATE: 25-APR-1997
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/851,843
 FILING DATE: 06-MAY-1997
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/854,050
 FILING DATE: 09-MAY-1997
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/911,312
 FILING DATE: 14-AUG-1997
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/912,951
 FILING DATE: 14-AUG-1997
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/915,503
 FILING DATE: 14-AUG-1997
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: WO PCT/US97/17618
 FILING DATE: 01-OCT-1997
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: WO PCT/US97/17885
 FILING DATE: 01-OCT-1997
 ATTORNEY/AGENT INFORMATION:
 NAME: Apple, Randolph Ted
 REGISTRATION NUMBER: 36,429
 REFERENCE/DOCKET NUMBER: 015389-002610US
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (415) 576-0300
 INFORMATION FOR SEQ ID NO: 110:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1031 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-974-549A-110

Query Match 4.0%; Score 100; DB 4; Length 1031;
 Best Local Similarity 21.5%; Pred. No. 0.41;
 Matches 106; Conservative 71; Mismatches 179; Indels 136; Gaps 29;

QY 18 LAMWIFTFQNSTKVMASLNLSLHYNNSTKSLFKPTPLSL-KPLETELRT---K 74
 Db 30 LYSWI-----QKVICRNSQS-HIKDLEDIKIFAQTNIVATPRDYNEEDPKVIARK 80
 QY 74 EI-----IEKLDQIIPPRFTHTVTTTSATHTATILNPRDYCRGDQ---HILLEVR 124
 Db 81 EVFSTGLMIELIDKCL-----VELLSSDVSVDROKLCFCGFCQKNGQLAKTHLLTA-- 131
 QY 125 DHLGRKQYGGDFLRARMSPALMAGASGVTFDFNNGTYLVSFTLFWEQVSLLLHP 184

Fri Oct 12 16:09:16 2001

us-09-729-454-1.rai

Page 5

[illegible]

RESULT 6
US-08-879-561-10
Sequence 10 Application US/08879561
Patent No. 5817482

GENERAL INFORMATION:

APPLICANT: Bandman, Olga
APPLICANT: Hillman, Jennifer L.
APPLICANT: Hawkins, Phillip R.
APPLICANT: Guegler, Karl J.
APPLICANT: Corley, Neil C.
TITLE OF INVENTION: DISEASE RELATED NUCLEOTIDE KINASES
NUMBER OF SEQUENCES: 12

CORRESPONDENCE ADDRESS:

ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/879,561
FILING DATE: Herewith
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PR-0125 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
TELEX:

INFORMATION FOR SEQ ID NO: 10:

SEQUENCE CHARACTERISTICS:
LENGTH: 610 amino acids
TYPE: amino acid
COMPOSITION: single
Topology: linear

; IMMEDIATE SOURCE:
 ; LIBRARY: GenBank
 ; CLONE: 705385
 US-08-879-561-10

Query Match 3.88; Score 95; DB 2; Length 610;
Best Local Similarity 24.98; Pred. No. 0.56;
Matches 47; Conservative 28; Mismatches 60; Indels 54; Gaps 11;

```

QY 215 VSSEGL-----ILNINAEGLQYDNRQGGFY-----CYRQHMPCALTH 250
Db 130 LHEQGLLEFFECFDPDPLDVC---EQRVKGLYKRAKAGIKGFTGIDQYESPAPEIQ 186
QY 257 MTSKKKYSYLSQKSLPERSNGVVEIMEKFNITSVSKCNLTSLKSDYHSGKLOHOLA- 316
Db 187 LKAGKSLIDECQVGVSLQKNGV-----VFESVNIIVKELVPSGC-LEHAAE 235
QY 316 -VDDL---RNINQW-----QKYCVPLIGSMYSYKMEYLTR-----AIDRTGKENT 360
Db 236 IYDFTPEMELTKLTDQWQVLSQGNATPLTGFN---REBEYLSQHGHCILD--GGVINO 289
QY 361 VVVISLQGH 369
Db 290 STIPIVLPHV 298

```

RESULT 7
 US-08-755-587-184
 : Sequence 184, Application US/08755587
 : Patent No. 6045997
 :
 : GENERAL INFORMATION:
 : APPLICANT: Futreal, Phillip A
 : APPLICANT: Wooster, Richard F
 : APPLICANT: Ashworth, Alan
 : APPLICANT: Stratton, Michael R
 : TITLE OF INVENTION: Materials and methods relating to the
 : identification and sequencing of the BRA2 cancer
 : TITLE OF INVENTION: susceptibility gene and uses thereof.
 : NUMBER OF SEQUENCES: 222
 :
 : CORRESPONDENCE ADDRESSES:
 : ADDRESSEE: Bell Seltzer Park & Gibson
 : STREET: 310 UCB Plaza, 3605 Glenwood Avenue, PO Drawer 31107
 : CITY: Raleigh
 : STATE: NC
 : COUNTRY: USA
 :
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: floppy disk
 : COMPUTER: IBM PC compatible
 : OPERATING SYSTEM: PC-DOS/MS-DOS
 : SOFTWARE: Patent Release #1.0, Version #1.25 (EPO)
 :
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/08/755,587
 : FILING DATE: 25-NOV-1996
 :
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: GB 9523959.6
 : FILING DATE: 23-NOV-1995
 :
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: GB 9525555.0
 : FILING DATE: 14-DEC-1995
 :
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: GB 9617961.9
 : FILING DATE: 28-AUG-1996
 :
 : ATTORNEY/AGENT INFORMATION:
 : NAME: Kenneth D Sibley
 : REGISTRATION NUMBER: 31,665
 : REFERENCE/DOCKET NUMBER: 5405-135
 : INFORMATION FOR SEQ ID NO.: 184:
 : SEQUENCE CHARACTERISTICS:
 : LENGTH: 1579 amino acids
 : TYPE: amino acid
 : TOPOLOGY: linear
 :
 : US-08-755-587-184

Query Match 3.8%; Score 94.5; DB 3; Length 1579;
 Best Local Similarity 19.6%; Pred. No. 3.1; Mismatches 169; Indels 181; Gaps 23;

Matches 102; Conservative 69; Mismatches 169; Indels 181; Gaps 23;

QY 48 NSTKSLFP--KTPPLSLKPLTE-----TELRIKE-IIEKIDQIPRPFTVHTTTS 96
 DB 431 NSNNLTPSOAKETELSTLESGQFEFTQFRPSYILQKSTFEVPMOMTILKTSE 490
 QY 97 ATNSTATILNPDYTCRQDOLHILEVVD-HLGRKQYGGDFLRAMGSPALM---AGA 151
 DB 491 E-----CRDADLHVMNAPSIVYDSSKQFECTVEIRKRAGLLKMDCKSA 536
 QY 152 SGKVTDFN-----NCTIVSFTLFEGOVYSLSLLIHPSGVSA-----LMSARNO 197
 DB 537 SGVLIDENEVGFRGYSAHGTKLNTSTALQAVKLFSDIENISETSAEVHPISLSSSK 596
 QY 198 GYDRVITFGQFVN---GTSQVHSECGLLNTNMLC-----OYLDNRQDE----- 239
 DB 597 CHDSVSMKEIENHNDKTSKNNKQOLLQNNIENTGTVEITEYKNTKRTENEDNKT 656
 QY 240 -----GFYVR-----POHPCALTHMYSK----- 260
 DB 657 AASRNSHNLFFDGSDDSKNDVTYVHKDETDLFTDOHNLCLKSGQFMKEGNTQIKEDLS 716
 QY 261 -----NKKVSYLSKQEKSL--FERSNVGVEIMEKFNITSYVSKCNTLK 300
 DB 717 DLTFVANAQOACHGNTSNKEQLTAKTEONIKOFESOTFFQTASGKN-ISVAKSEFVK 775
 QY 301 SV-----DLHESGLOHOLAVDLNR-----IN 323
 DB 776 IVNFQKPEELH-NESLNSLHSDIRKNNKMDLISEETDIYKHLIKESPVGTGNOQLT 834
 QY 324 ICNQ-----KYCP-LIGSMYTSVKEMEYLTRAIDRTG-----EKNTVYISLQGHFPP 372
 DB 835 FQGOERDEKIKEPTLLGFTLSGKKVYIAKESLDVKNLFDENKQGGSEITSPSNOW-- 892
 QY 373 FPIQVIRRALNVKAIQHLILRSPDVIYIKTENIREMYN 413
 DB 893 -----AKTLKYREACKDLEL-ACEITEITIAAPKCKEMON 925

RESULT 8
 US-08-804-439A-17
 ; Sequence 17, Application US/08804439A
 ; Patent No. 6015565
 ; GENERAL INFORMATION:
 ; APPLICANT: Rose, Timothy M.
 ; APPLICANT: Bosch, Martin L.
 ; APPLICANT: Strand, Kurt
 ; TITLE OF INVENTION: GLYCOPROTEIN B OF THE RFHV/KSHV
 ; NUMBER OF SEQUENCES: 113
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Fish & Richardson P.C.
 ; STREET: 4225 Executive Square, Ste 1400
 ; CITY: La Jolla
 ; STATE: CA
 ; COUNTRY: USA
 ; ZIP: 92037
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/804,439A
 ; FILING DATE: February 21, 1997
 ; CLASSIFICATION: 424
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Hallie, Lisa A.
 ; REGISTRATION NUMBER: 38,347

REFERENCE/DOCKET NUMBER: 09176/004001
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (619) 678-5070
 TELEFAX: (619) 678-5099
 TELEX:
 ; INFORMATION FOR SEQ ID NO: 17:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 849 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-804-439A-17

Query Match 3.7%; Score 92; DB 3; Length 849;
 Best Local Similarity 18.7%; Pred. No. 2;
 Matches 86; Conservative 82; Mismatches 172; Indels 120; Gaps 20;

QY 25 TYFQNSTKWSA---LNSIS--LHYNN-----STKSLFPKP--LISLKPTELEUR 71
 DB 353 TEVENFTSTSCVSDQIDINKTISEYIOKLNSYVASKQYFXTDGNLYLIMQPLEHPE-- 410
 QY 72 IKETIEKIDQIPRPFTVHTTTSATHTATILNPDYTCRQDOLHILEVNDHGRKK 131
 DB 411 -----IEDIDSDDEPTPAPKSTKRREKADNGNSTSEVSKSENPILT-----AVI 459
 QY 132 QYGGQFLFARMS---PALMACSKVTDENGTLYVSTFLFMEGOVLSLLIHPSGV 188
 DB 460 QPAYDKLTTSVNVYLELSRANCFQVAD-----TLKW-----YELSKVNTSTW 504
 QY 189 SALW-----SARNG-----YHNAVYQYVHNS 213
 DB 505 SAIVKRYPARARYGDAISTDCTIYDSSVNIHOSLRLOHDKTTGYSFRTYKFINSTL 564
 QY 214 OVHSCG---LILTNALCOYLNDKDEGFCYRPVHMCALTHMYSKRVSYLS 268
 DB 565 PLTDGQRRKEIILSTNTEIC---KDESEHYFVGEYI-----YYKN---YIF 608
 QY 269 KQKSLFERSNVGVEIMEKFNITSYVSKCNTL--KSYDLHESGLOHOLAVDLNRNINQW 726
 DB 609 EKKLNL-----SSIALDLFTIALNISTFENIDKTYELVSTKRLASSVFDIESPHEV 663
 QY 327 QKCYPLIG-----SMTYS---VKEMEYLTRAIDRTGSEKNTVYISLQGHFPP 375
 DB 664 NYTYSLAGIKKDLDTIDYNDRLOVOLDMDYADGDIG--RSVNVVSSVYTFESSIV 721
 QY 376 DVFTIRALNVKAIQHLILRSPDVIYIKTENIREMYN 415
 DB 722 TGFIKFTNPLGIFILILIGIIFLVVYLNRNSQFHDA 761

RESULT 9
 US-08-720-229-17
 ; Sequence 17, Application US/08720229
 ; Patent No. 6022542
 ; GENERAL INFORMATION:
 ; APPLICANT: Rose, Timothy M.
 ; APPLICANT: Bosch, Martin L.
 ; APPLICANT: Strand, Kurt
 ; TITLE OF INVENTION: GLYCOPROTEIN B OF THE RFHV/KSHV
 ; NUMBER OF SEQUENCES: 100
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Morrison & Foerster
 ; STREET: 755 Page Mill Road
 ; CITY: Palo Alto
 ; STATE: CA
 ; COUNTRY: USA
 ; ZIP: 94304-1018
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible

```

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/720,229
FILING DATE: 26-SEP-1996
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Schiff, J. Michael
REGISTRATION NUMBER: 40,253
REFERENCE/DOCKET NUMBER: 29938-20002.00
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 813-5600
TELEFAX: (415) 494-0742
FLEX: 706141
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 849 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-720-229-17

```

```

Query Match      3.7% Score 92; DB 3; Length 849;
Best Local Similarity 18.7% Pred. No. 2;
Matches 86; Conservative 82; Mismatches 172; Indels 120; Gaps 20;

```

```

QY 25 TVEONSTKVMAS---LNLSIS--LHWNN---STKSLPPTP---DISLKLPTETELR 71
DB 35 TVEWNTSTVSCVSDQINKTISEYIOKINSYVASGCTQYFRTDGNLILIQPLENHE-- 410
QY 72 IKEIIEKLDQOIPRPETHVNTTTSATHTATILNPRDTYCGDGLHILEVRDHLGRK 131
DB 411 ----IEDIDEDSDPEPTPAPKSTRKREADNCGNSTSEVSGSENPLIT-----AQI 459
QY 142 QVGGDFLRARMS---PALMAGASGKVTDFNNGTIVSFTLFMEQVSLSLILHPSGV 188
DB 460 QFAYDKLTTSVNVVLELSEAMCREQVRD-----TLWM-----YELSKNPTISW 504
QY 189 SALM-----SARNQ-----YDVIFTGQFVNGTS 213
DB 505 SATVGRVYAAHYVDALSVTDCIYVDOSSVNIHUSLRLQHDKTCYSRPVTFKFINSTD 564
QY 214 QVHSECG---LILTNALQVLDNRDQGFYCVARPOHPCALTHMSKKNKVSYS 268
DB 565 PLTGQLGRKEIILSNMNIETC---KDESEHYFTIGEVI-----YKKN---YIF 608
QY 269 KQKSLFPRSNVGEIEMKFNITSKCNLT--KSYDLHESKLOHOLAVDLDRNINIQM 326
DB 609 EKKLNL-----SSIALDIFIALNISFTLENDKTVELYSSTERKLASSYFIDESMREY 663
QY 327 QKCYCPILG-----SKTYS---VKEMEYLTRAIDRTGKKNVTYISLQGFRRPPI 375
DB 664 NYTYSLAGIKKDLDTIDYNDRLVQDLSDMADLDGIG--RSVVNVVSSVVTFFSSIV 721
QY 376 EYFIRRALNVHKAIOHLILRSPDTVIIKTENIREMYNDA 415
DB 722 TGFIKFTINPLGGIFTLLIIGGITLVVVLNRRNSQFMDA 761

```

```

RESULT 10
US-08-417-089-6
Sequence 6, Application US/08417089
Patent No. 6069298
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: METHODS AND AN ACETYL COA CARBOXYLASE GENE
TITLE OF INVENTION: FOR CONFERRING HERBICIDE TOLERANCE AND AN ALTERATION IN
NUMBER OF SEQUENCES: 11
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

```

```

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/417,089
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 2325 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-417-089-6

```

```

Query Match      3.6% Score 90.5; DB 3; Length 2325;
Best Local Similarity 18.8% Pred. No. 16;
Matches 99; Conservative 78; Mismatches 178; Indels 171; Gaps 23;

```

```

QY 63 KPLTELEIKETIEKLDQ-OIPRPETHVNTTTS-----ATHSTATILNPRDTY----- 111
DB 659 KVTYTNAAVSEVSYLTNGHISLVSTVNNINIBSKYTIETFTGCHGSYRLRMN 718
QY 112 -----CRGDQ-----HILEVDHILGRKQYGGD--FLARMSPALMAGA 151
DB 719 DSTVEANVQSLDQGLLMQLDGNSHVITAEBAAGTRLDQDKTCLLDNDHDSKILLET 778
QY 152 SGQVYDF--NNGTIVSFTLFMEQV--SLSLILHPSGVSAKSNRNGT---DVT-- 203
DB 779 PCKLRLFLVADGAAHVADPYAEVEMKMCPLDSPASGVYHQMMSQEQALQADLIARL 838
QY 204 -----FTQGF-----VNGTSQVHSECGILINT----- 225
DB 839 DLDDPSAVKRAEPEDGIFPQMEPLPAVSSQYHKRYAASLNAARVYLAEGHININEVQDL 898
QY 226 -----NAEL--CQY-----LDNRDQ---GFTCVARPOHPCALTHM 257
DB 899 VCCLDNPEPLQWDELMSVLTARLRLNLSKSELEDYKEKLNFTYHGNDPFSKLLRDI 958
QY 258 YSKNKNVSYLSKQKSLFER-----SNVGEIEMKFNITSKCN 297
DB 959 IEEN--LSYGSEKKEKATNELEPLMLNLKSYGSESHAHFVYKSLPEEVLLELFS 1016
QY 298 TLKSYDLHESKLOHOLAVDLDRNINIQMOCYCPILGSMTSYKEMETLRAIDTQGE 357
DB 1017 GLOS-DVLET--LHSHSKDQKQVDT-----VLSHQVYR 1048
QY 358 KNTVIVISLQGF--RFP---IDVIFIRRALNVHKAIOHLILRSPDTVIIKTENIREMYN 413
DB 1049 NNAKLVTALMEKLEVYRPGGYRDLVYRSSLNHKRYKALAKASELLEQTKSELRA--S 1106
QY 414 DAERPSDFHYIOYLKIDIFQDL-----SVSLIDAMDTI 449
DB 1107 VARSLSDLMHKMGKMSIKDMEDVSAFLPYEDALISLDPIYSDRTV 1152

```

```

RESULT 11
US-08-695-651-6
Sequence 6, Application US/08695651
Patent No. 6146867
GENERAL INFORMATION:
APPLICANT: Geogenbach, B. G.
APPLICANT: Somers, D. A.
APPLICANT: Myae, D. L.
APPLICANT: Gronwald, J. W.
APPLICANT: Egli, M. A.
APPLICANT: Lutz, S. M.
TITLE OF INVENTION: METHOD FOR PRODUCING PLANT ACETYL COA
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESS: Schwegman, Lundberg, Woessner & Kluth, P.A.
STREET: P.O. Box 2938
CITY: Minneapolis

```

STATE: MN
 COUNTRY: USA
 ZIP: 55402
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette
 COMPUTER: IBM Compatible
 OPERATING SYSTEM: DOS
 SOFTWARE: FASTSEQ Version 2.0
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/695,651
 FILING DATE: 12-AUG-1996
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/417089
 FILING DATE: 05-APR-1995
 APPLICATION NUMBER: 08/014326
 FILING DATE: 05-FEB-1993
 APPLICATION NUMBER: 07/917462
 FILING DATE: 21-JUL-1992
 APPLICATION NUMBER: 07/538674
 FILING DATE: 18-JUN-1990
 ATTORNEY/AGENT INFORMATION:
 NAME: Noesner, Warren D
 REGISTRATION NUMBER: 30,440
 REFERENCE/DOCKET NUMBER: 600,318US3
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 612-373-6900
 TELEFAX: 612-339-3061
 TELEX:
 INFORMATION FOR SEQ ID NO: 6:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2325 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 FRAGMENT TYPE: N-terminal
 US-08-695-651-6

Query Match 3.6%; Score 90.5; DB 4; Length 2325;
 Best Local Similarity 18.8%; Pred. No. 16;

Matches 99; Conservative 78; Mismatches 178; Indels 171; Gaps 23;

QY 63 KPLTELEIRKEIIEKLDQ-QIPPRPTHVTTTS-----ATHSTATILNPRDY----- 111
 DB 659 KVTNNATVSEYVSYLTGKHIPKHSILVNSTVINLIEGSKYTIETVKGHSGYLRNN 718
 QY 112 -----CRGDQ-----HILLEVRDHLGRKOYGD--FLPARRSSPALMAGA 151
 DB 719 DSTVEANVQSLCDGGLMDQNSHVIYAEENGSTRQIDGKTCLLQNDHDSKLLAET 778
 QY 152 SGNVTD--NNGTYLVSTFLFWEGOV-SLSLLIHPSGVSALMSARNQY--DVI-- 203
 DB 779 PCKLRLVADGAHVADVPYAEVEVMKMCPLSPASGVTHCMSSGQALQAGDILARL 838
 QY 204 -----FTGQF-----VNGISOVHSEGLILNT----- 225
 DB 839 DDDPSAKVRAEPFDGIFQMLFPYAVSSOVHRKYVASLNAARVLGYEHNINEVVDL 898
 QY 226 -----NAEL--COY-----LDNRDQ--GFYCVRPOHPCALTHM 257
 DB 899 VCCLDNPELPLOMDLSVLTALRLPRLKSELEDKYKEYKLNLYHOKNDPFSKLLADI 958
 QY 258 YSKNKVSYLSKQSKSLFER-----SNVGVLEMEFNTISVSKCN 297
 DB 959 IREN--LSYSEKEKATNERLVEPLMLMLSYEGGSHAFVVKSLFEYLTVELEFSD 1016
 QY 298 TLKSYDLHESGKLQHOVLADNRININQOWKYCYPLIGSMYTSVKEKEXYLTRAIDRTGE 357
 DB 1017 GIGD-VIET--LRHOSKDLQKVVDI-----VLSHGQVR 1048
 QY 358 KNTYVVISLQGHF-RFPF---IDVFIRRALNVKKAIOHLLRLSPDVTWIIKTENIREKYN 413

DB 1049 NKAIVTALMEKLYPNPBGYRDLVTRSSLNHRKYKLLAKASLEIEIKLSLFA--S 1166
 QY 414 DAEFSDPHGYIOYLIIRKIDFODL-----SVSTIDAWIT 449
 DB 1107 VARSLSDLQMKHSEKSIKDMNEDLVAPLFEADLISLFQYSRRTV 1152

RESULT 12

US-08-930-285-6
 Sequence 6 Application US/08930285
 Patent No. 6222099

GENERAL INFORMATION:

APPLICANT: Regents of the University of Minnesota, et al.
 TITLE OF INVENTION: TRANSGENIC PLANTS EXPRESSING ACETYL COA CARBOXYLASE II
 NUMBER OF SEQUENCES: 26
 CORRESPONDENCE ADDRESS:

ADDRESSEE: Schwegman, Lundberg, Woessner & Kluth, P. A.
 STREET: P. O. Box 2938
 CITY: Minneapolis

STATE: MN
 COUNTRY: USA

ZIP: 55402
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible
 OPERATING SYSTEM: DOS

SOFTWARE: FASTSEQ Version 1.5
 CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/930,285
 FILING DATE: 13-APR-1998
 CLASSIFICATION: 800

PRIOR APPLICATION DATA:
 APPLICATION NUMBER: PCT/US96/04625

FILING DATE: 04-APR-1996
 ATTORNEY/AGENT INFORMATION:

NAME: Embretson, Janet E.
 REGISTRATION NUMBER: 39,665

REFERENCE/DOCKET NUMBER: 600,318US4
 TELECOMMUNICATION INFORMATION:

TELEPHONE: 612-339-0331
 TELEFAX: 612-339-3061

TELEX:
 INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:
 LENGTH: 2325 amino acids

TYPE: amino acid
 STRANDEDNESS: single

TOPOLOGY: linear
 MOLECULE TYPE: protein

HYPOTHETICAL: NO
 ANTI-SENSE: NO

FRAGMENT TYPE: N-terminal and C-terminal (full length protein)
 ORIGINAL SOURCE:

US-08-930-285-6

Query Match 3.6%; Score 90.5; DB 4; Length 425;
 Best Local Similarity 18.8%; Pred. No. 16;

Matches 99; Conservative 78; Mismatches 178; Indels 171; Gaps 23;

QY 63 KPLTELEIRKEIIEKLDQ-QIPPRPTHVTTTS-----ATHSTATILNPRDY----- 111
 DB 659 KVTNNATVSEYVSYLTGKHIPKHSILVNSTVINLIEGSKYTIETVKGHSGYLRNN 718
 QY 112 -----CRGDQ-----HILLEVRDHLGRKOYGD--FLPARRSSPALMAGA 151
 DB 719 DSTVEANVQSLCDGGLMDQNSHVIYAEENGSTRQIDGKTCLLQNDHDSKLLAET 778
 QY 152 SGNVTD--NNGTYLVSTFLFWEGOV-SLSLLIHPSGVSALMSARNQY--DVI-- 204
 DB 779 PCKLRLVADGAHVADVPYAEVEVMKMCPLSPASGVTHCMSSGQALQAGDILARL 838

```

QY 204 -----FCGOF-----VNGTSOYHSECCLLINT----- 225
Db 839 DLDPSAVKRAEPFDGTFPOMELPAVSVQVHKRYTASLMAARMLYAEHNINEVYQDI 898
QY 226 -----NAEL--COY-----LDNRDQ--GFCVBPQHPQALTHM 257
Db 899 VCCLDNPELPLOWDELMVSLATRLPRLNKSELEDKYEKLYKFNEDPSPKLLROI 958
QY 258 YSKNKKVSYLSKOKSLPER-----SANGVEIMEKNITISYKCN 297
Db 959 IFEEN--LTSYSEKAKATNERLVEPLMLNKSEKESNAHNVKSLSEELTVEELFSD 1016
QY 298 TKSVDLHESGKLOHOLAVLDNRNINIQOKYCPPLGSMYSKEMEYLTRAIDRTGE 357
Db 1017 GLOS-DVLET--LRHOSKDLQKVVDI-----VLSHQVGR 1048
QY 358 KNTYVVISLQGHF-RFP--IDVFIRRALNVKALQIHLHLSPTQVYIKTINIREMN 413
Db 1049 NKAKLVATLMEKLVYVNPQGYKDLVFRSSLNKRYKYLKALASELEQTKSELNA--S 1106
QY 414 DAERFSDFHGYIOYLLIKDIFODL-----SVSIIDANDIT1 449
Db 1107 VARSISDLGMHKGEMSIKDMEDVSAPLPEYDALISDFYSDRTV 1152

RESULT 13
US-08-376-362A-20
: Sequence 20, Application US/08376362A
: Patent No. 5693756
: GENERAL INFORMATION:
: APPLICANT: LI, Xiao-Jiang
: APPLICANT: Blackshaw, Seth
: TITLE OF INVENTION: AMIOTRIDE-SENSITIVE SODIUM CHANNEL AND
: TITLE OF INVENTION: METHOD OF IDENTIFYING SUBSTANCES WHICH STIMULATE OR BLOCK
: TITLE OF INVENTION: SALT TASTE PERCEPTION
: NUMBER OF SEQUENCES: 20
: CORRESPONDENCE ADDRESS:
: ADDRESSER: Hammer & Alletrettl, LTD
: STREET: 1001 G Street, N.W., Eleventh Floor
: CITY: Washington
: STATE: D.C.
: COUNTRY: U.S.A.
: ZIP: 20001-4597
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/376,362A
: FILING DATE: 23-JAN-1995
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Karen A., Sarah
: REGISTRATION NUMBER: 32,141
: REFERENCE/DOCKET NUMBER: 01107, 48125
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 202 508-9100
: TELEFAX: 202 508-9299
: INFORMATION FOR SEQ ID NO: 20:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 698 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: Molecule type: protein
: HYDROTICAL: NO
: ANTI-SENSE: NO
: ORIGINAL SOURCE:
: ORGANISM: Rattus rattus
: US-08-376-362A-20

```

```

Query Match 3.6%; Score 90; DB 1; Length 698;
Best Local Similarity 18.9%; Pred. No. 2.3;
Matches 102; Conservative 47; Mismatches 147; Indels 244; Gaps 25;

QY 23 IFVFNSTKYKMSALNLSLH-----YWN----- 47
Db 83 LQFPCNNTTHGAILVCSKHNRKTLFMAVLMCTFCGMWOPALFEELYSTPVLN 142
QY 48 ---NSTKSLFPKTPPLISLKPTELRIKELIKDQ----- 81
Db 143 INLNSDKLVFPAYVCTLPNRYTE--IKEELELDRIEQLTFLQYKNSVTRQAGAR 200
QY 82 -----QIPRPTHVNTTSTATSTALINR----- 108
Db 201 RRSRDILGAFPHLDRLTTPPYIS--GRTANSSSVYRDNPPQVCKMKIGFOLCQ 258
QY 109 -----DTYRGDQHLILEVRDHLGRROYGDFLRARMSPALMAGASG----- 153
Db 259 NKSDFCYQYVSSG---VDAYREMY--RFHYINILSLDPTSPALPEEALGNFIFCRF 311
QY 154 -----TSQV--HS-----ECGLINTNAELQYLDNRDQES--FY- 242
Db 312 NQAPQNAKSKFHHPMYGNCTFNDKNSNLMASMPGVNNG--LSLTRTEQNDPFP 368
QY 179 LL-----LHPSE-----GVALMSARNOGYDRAVIFTGQF-----VNG- 211
Db 369 LLSTVIGARVAVHGOPEAFMDGQFNLRPQVETSIKMKALDSL--GGNYCDCTENG 426
QY 212 -----TSQV--HS-----ECGLINTNAELQYLDNRDQES--FY- 242
Db 427 DVYVKNLYPKYTYQVCHISCFQEMIKKGCAGVITFKPQGVFCYRKQSSWGYCYK 486
QY 243 -----CYRPQHPQCALTHM-----YSKNNKVSYSKOKSLFERSNV--GVEM 285
Db 487 LOGAFSLDSLGCSKRCRKCPCSVIYKLSAGYSRMPV-----KSDWIFEMLSLQNNYIN 542
QY 286 EKFNITSYKCNLT-LKSYDLHESGKLOHOLAVLDNRNINIQOKYCPPLGSMYSKEM 344
Db 543 NKRN--GVAKLNTFFKELNKTNSSESPYVMSLSLNSGQMSLN-----FGSVLSYVEM 596

RESULT 14
US-08-530-950-13
: Sequence 13, Application US/08530950
: Patent No. 5736381
: GENERAL INFORMATION:
: APPLICANT: Davis, Roger J.
: APPLICANT: Raingeaud, Joel
: APPLICANT: Gupta, Shaahl
: APPLICANT: DeJaird, Benoit
: TITLE OF INVENTION: CYTOKINE-, STRESS-, AND
: TITLE OF INVENTION: ONCOPROTEIN-ACTIVATED HUMAN PROTEIN KINASE
: TITLE OF INVENTION: KINASES
: NUMBER OF SEQUENCES: 16
: CORRESPONDENCE ADDRESS:
: ADDRESSER: Fish & Richardson P.C.
: STREET: 225 Franklin Street
: CITY: Boston
: STATE: MA
: COUNTRY: USA
: ZIP: 02110-2804
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/530,950
: FILING DATE: 19-SEP-1995
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Fasse, J. Peter
: REGISTRATION NUMBER: 32,983

```

REFERENCE/DOCKET NUMBER: 07917/010001
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 617/542-5070
 TELEFAX: 617/542-8906
 TELEX: 200154
 INFORMATION FOR SEQ ID NO: 13:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 668 amino acids
 TYPE: amino acid
 STRANDEDNESS: not relevant
 TOPOLOGY: linear
 US-09-530-950-13

Query Match 3.6%; Score 89.5; DB 1; Length 668;
 Best Local Similarity 18.8%; Pred. No. 2.4; Mismatches 134; Indels 109; Gaps 14;
 Matches 72; Conservative 69;

QY 60 ISLKLPLETELRLKEIIEKLDQIIPRPFT--HVNTTTS-----AT 98
 DB 230 MSLKMPRT-----AQPQOFAPSPSNKKHIEITLSKVEGKRNPGSLINGVQT 261
 QY 99 HSTATILNPRDTCRGDQHLILEVRDLGRKQYGGDFLRARMSPALMAGASGVYDF 158
 DB 282 STSSSTEGPHDT-----VGTTPRTG---NSNNSNGSGSGGGGLFANF 321
 QY 159 NNGTYLVSFTLFWEGOVSLILLHPSEGSALMSARNQGYDRVITGQFVNGTSQVHSE 218
 DB 322 SKYVDIKSGSLNFAKGLSSSGIDFSGSSSRITLDELFLDELGHGNGNYSKYLHP 381
 QY 219 CGLLINT-----NAELCOYLDNRDQ-----EGFYCVRPQIMVYA 252
 DB 382 TNVIMATKEVRLDELDAKFRQILMELEVLHKNSPYIVDFGAFIEGAVYMWCMEMOGG 441
 QY 253 ALTHMSKKNKVSYSKOE-----KSLFRSNVGEIEMKFNITSVSKCNTLK 300
 DB 442 SLDKTYDESSSEIGIDEPQLAFIANAVIHGELKEQHNIIHROVKFTNLLCSANGTVK 501
 QY 301 SYDLHESGKLQHLAVDLDRNINIQWOKICYR-LIGSM-----TYSVKEEYTLIAIDRT 354
 DB 502 LCDFGVSGNLVASLA-----KTINIGQSYMAPERIKSLNPDRAITYV-QSDIWSLGL 552
 QY 355 GGEKNIVIVISLGHFRFPPIDV 378
 DB 553 -----SILEMALGRY--PYPPETY 569

RESULT 15
 US-09-149-879-13
 Sequence 13, Application US/09149879
 Patent No. 6174676
 GENERAL INFORMATION:
 APPLICANT: Davis, Roger J.
 APPLICANT: Raingeaud, Joel
 APPLICANT: Gupta, Shashi
 APPLICANT: Derliard, Benoit
 TITLE OF INVENTION: CYTOKINE, STRESS-, AND
 TITLE OF INVENTION: ONCOPROTEIN-ACTIVATED HUMAN PROTEIN KINASE
 NUMBER OF SEQUENCES: 16
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Fish & Richardson P.C.
 STREET: 225 Franklin Street
 CITY: Boston
 STATE: MA
 COUNTRY: USA
 ZIP: 02110-2804
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/149,879
 FILING DATE:
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/530,950
 FILING DATE:
 ATTORNEY/AGENT INFORMATION:
 NAME: Fasse, J. Peter
 REGISTRATION NUMBER: 32,983
 REFERENCE/DOCKET NUMBER: 07917/010001
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 617/542-5070
 TELEFAX: 617/542-8906
 TELEX: 200154
 INFORMATION FOR SEQ ID NO: 13:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 668 amino acids
 TYPE: amino acid
 STRANDEDNESS: not relevant
 TOPOLOGY: linear
 US-09-149-879-13

Query Match 3.6%; Score 89.5; DB 4; Length 668;
 Best Local Similarity 18.8%; Pred. No. 2.4; Mismatches 134; Indels 109; Gaps 14;
 Matches 72; Conservative 69;

QY 60 ISLKLPLETELRLKEIIEKLDQIIPRPFT--HVNTTTS-----AT 98
 DB 230 MSLKMPRT-----AQPQOFAPSPSNKKHIEITLSKVEGKRNPGSLINGVQT 261
 QY 99 HSTATILNPRDTCRGDQHLILEVRDLGRKQYGGDFLRARMSPALMAGASGVYDF 158
 DB 282 STSSSTEGPHDT-----VGTTPRTG---NSNNSNGSGSGGGGLFANF 321
 QY 159 NNGTYLVSFTLFWEGOVSLILLHPSEGSALMSARNQGYDRVITGQFVNGTSQVHSE 218
 DB 322 SKYVDIKSGSLNFAKGLSSSGIDFSGSSSRITLDELFLDELGHGNGNYSKYLHP 381
 QY 219 CGLLINT-----NAELCOYLDNRDQ-----EGFYCVRPQIMVYA 252
 DB 382 TNVIMATKEVRLDELDAKFRQILMELEVLHKNSPYIVDFGAFIEGAVYMWCMEMOGG 441
 QY 253 ALTHMSKKNKVSYSKOE-----KSLFRSNVGEIEMKFNITSVSKCNTLK 300
 DB 442 SLDKTYDESSSEIGIDEPQLAFIANAVIHGELKEQHNIIHROVKFTNLLCSANGTVK 501
 QY 301 SYDLHESGKLQHLAVDLDRNINIQWOKICYR-LIGSM-----TYSVKEEYTLIAIDRT 354
 DB 502 LCDFGVSGNLVASLA-----KTINIGQSYMAPERIKSLNPDRAITYV-QSDIWSLGL 552
 QY 355 GGEKNIVIVISLGHFRFPPIDV 378
 DB 553 -----SILEMALGRY--PYPPETY 569

Search completed: October 12, 2001, 16:03:51
 Job time: 74 sec

Fri Oct 12 16:09:16 2001

us-09-729-454-1.rai

Page 11

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

QM protein - protein search, using sw model

Run on: October 12, 2001, 16:02:37 ; Search time 42.34 Seconds

(without alignments)
1484.291 Million cell updates/sec

Title: US-09-729-454-1

Perfect score: 2495

Sequence: 1 MKISMINKSLALFLIAS.....VHPQHVGNQINILNYIC 475

Scoring table: BLOSUM62

Search: Gapop 10.0 , Gapext 0.5

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Database: 1

SPREMBL_16:*

1: sp-archaea:*

2: sp-bacteria:*

3: sp-fungi:*

4: sp-human:*

5: sp-invertebrate:*

6: sp-mammal:*

7: sp-misc:*

8: sp-organelle:*

9: sp-phage:*

10: sp-plant:*

11: sp-rodent:*

12: sp-unclassified:*

13: sp-vertebrate:*

14: sp-virus:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	340.5	13.7	160	4	Q9NXP5	Q9NXP5 homo sapien
2	129.5	5.2	744	11	Q9ESN6	Q9ESN6 mus musculu
3	127.5	5.1	792	4	O60272	O60272 homo sapien
4	119.5	4.8	3603	2	P94459	P94459 bacillus su
5	112.5	4.5	4060	14	Q9IH28	Q9IH28 gill-assoct
6	109	4.4	688	5	O9TYZ7	O9TYZ7 caenorhabd
7	108.5	4.4	1387	5	O9GZ76	O9GZ76 plasmodium
8	106	4.3	2571	2	O87704	O87704 bacillus su
9	104.5	4.2	3262	4	O13788	O13788 homo sapien
10	104	4.2	728	4	O13779	O13779 homo sapien
11	102	4.1	800	5	O43988	O43988 dictyostell
12	99.5	4.0	1507	6	O9V873	O9V873 drosophila
13	98.5	4.0	795	10	O9STU3	O9STU3 canis famli
14	98.5	4.0	1004	5	O96167	O96167 plasmodium
15	98.5	3.9	470	3	O12049	O12049 saccharomyc
16	98	3.9	657	2	O9PNN6	O9PNN6 campylobact
17	98	3.9	867	13	O9W6C5	O9W6C5 xenopus lae
18	98	3.9	4699	3	O9HFB8	O9HFB8 schizosacch
19	98	3.9				

20	97.5	3.9	953	4	Q9Y239	Q9Y239 homo sapien
21	97	3.9	538	14	O83653	O83653 mumps virus
22	97	3.9	897	10	O41865	O41865 zea mays (m
23	97	3.9	1366	2	O54165	O54165 shigella fl
24	97	3.9	1802	3	O96170	O96170 plasmodium
25	96.5	3.9	1245	3	O03868	O03868 saccharomyc
26	96	3.9	538	14	O9JAE8	O9JAE8 mumps virus
27	96	3.9	952	5	O9VX01	O9VX01 drosophila
28	96	3.9	2692	5	O21547	O21547 caenorhabd
29	95.5	3.8	1285	14	O9DHP2	O9DHP2 yaba-like d
30	95.5	3.8	1685	10	O43248	O43248 zea mays (m
31	95	3.8	610	5	O27128	O27128 urechis cau
32	95	3.8	851	14	O9V584	O9V584 fowlpox vir
33	95	3.8	1021	5	O15733	O15733 dictyostell
34	94.5	3.8	740	10	O9SVH0	O9SVH0 arabidopsi
35	94.5	3.8	803	2	O82879	O82879 streptococc
36	94.5	3.8	2151	14	O9DY39	O9DY39 hantavirus
37	94.5	3.8	538	14	O9WAT0	O9WAT0 mumps virus
38	94	3.8	971	2	O51579	O51579 borrelia bu
39	94	3.8	1286	14	O90903	O90903 Shope fibro
40	94	3.8	3495	2	O30380	O30380 bacillus su
41	93.5	3.8	650	3	O9U085	O9U085 schizosacch
42	93.5	3.8	684	4	O9UXF6	O9UXF6 homo sapien
43	93.5	3.8	1209	5	P91581	P91581 cloa intest
44	93.5	3.8	1311	5	O9TYW6	O9TYW6 caenorhabd
45	93.5	3.8	2894	1	O58791	O58791 methanococc

ALIGNMENTS

RESULT	ID	Q9NXP5	PRELIMINARY	PRT	160 AA.
1	Q9NXP5				
AC	Q9NXP5				
DT	01-OCT-2000 (TREMblrel. 15, Created)				
DT	01-OCT-2000 (TREMblrel. 15, Last sequence update)				
DT	01-OCT-2000 (TREMblrel. 15, Last annotation update)				
DE	CDNA FLJ20127 FIS, CLONE COL06176.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
OX	NCBI_Taxid=9606;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=COLON;				
RA	Kawabata A., Hixji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,				
RA	Ohtani T., Ota T., Suzuki Y., Obayashi M., Nishi T., Shibahara T.,				
RA	Tanaka T., Nakamura Y., Isogai T., Sugano S.,				
RT	"NEDO human cDNA sequencing project."				
RL	Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.				
DR	EMBL: AK000134; BAA0965.1; "				
SO	SEQUENCE 160 AA; 18466 MW; 113114CA43912208 CRC64;				

Query Match	13.7%	Score 340.5;	DB 4;	Length 160;
Best Local Similarity	51.0%;	Pred. No. 4.1e-20;		
Matches	73;	Conservative	0;	Mismatches 1;
				Indels 69;
				Gaps 1;
QY	285	MEKFNFTSYSKC-----		296
DB	1	MEKFNFTSYSKC-----		296
QY	297	-----NLTKSVLDHSGKLOHQLAVDLDRNINQOKCTPLIG 335		
DB	61	LITLMDSTIRQMEHYRKASINILKSVLDHSGKLOHQLAVDLDRNINQOKCTPLIG 120		
QY	336	SMYTSYKENEYLTIRADRTGKK 358		
DB	121	SMYTSYKENEYLTIRADRTGKK 143		
RESULT	2			

Q9ESN6
ID Q9ESN6 PRELIMINARY; PRT: 744 AA.
AC Q9ESN6:
DT 01-MAR-2001 (TEMBLrel. 16, Created)
DT 01-MAR-2001 (TEMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TEMBLrel. 16, Last annotation update)
DE NEURAL ACTIVITY-RELATED RING FINGER PROTEIN.
GN NARF.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ICR MICE;
RA Ohkawa N.;
RT *cDNA sequence encoding Neural Activity-Related Ring Finger Protein
RT (NARF).
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB043550; BAB17634.1; -
SQ SEQUENCE 744 AA; 81445 MW; 80336DA2EFE3E5FD CRC64;

Query Match 5.2%; Score 129.5; DB 11; Length 744;
Best Local Similarity 22.1%; Pred. No. 0.051;
Matches 89; Conservative 43; Mismatches 133; Indels 137; Gaps 18;

QY 66 TETELRI--KEIEKL-----DOQIPRP-----FTHYNTTSA 97
DB 266 TETELVLYKKQMSKLNELADQDFLHPRENDQDIFVETBGLKSHINLCTILTINAVA 325
QY 98 THSTATILNPDIYCGDQHLILEVRDHLRRKQYGGDFLRARMSPALMAGASKVTD 157
DB 326 SETVATGEGLEKQITII-GQPMSTVITTKDKDELCKTGNAYLTAEISTPD-GSVADEGIID 383
QY 158 FNNGTIVLSFTLFMEQVSLSL-----IHP-SEGVSLMGNARND 197
DB 384 NKGTYEELTYQKEDFTLSRLYDQIHRSFPFKLYIRSDVSPFTTEGKKRVKSPGS 443
QY 198 GY-----DRVIF-----TGQFVNGTSVHSECG--LIL 223
DB 444 GHVKQKAVKRPASMYSTGKRKENPIEDDLIFRVGTGRNKGFTNLQGVAASTSGILIA 503
QY 224 NTNAELCYLDNRDQ-EGFYCVRPQ-----HMPCAALTMNS-----KNKAVSYLSNQ 270
DB 504 DSNNGCVQIFSDNGQFKSRFGIRGRSPQQLDRPTGYAVHPSDIIILADYDNKWSIFSSD 563
QY 271 EKSLFERNVGVETMEKENTISVSKNTLKSVDLHESGKLQHLAVDLDRN--INIQWCK 328
DB 564 GKF---KTKIG-----SGKLMPKGVSVDRNGHIIYVDRK 595
QY 329 YCYPLIGSMYSVKEMEYLTRAIDRTGGEKNTVIYISLQGF 370
DB 596 ACCVFIFQPNKGI-----VTRFGSRNGNDQFA-----GPHF 627
PRT: 792 AA.
ID 060272 PRELIMINARY; PRT: 792 AA.
AC 060272:
DT 01-AUG-1998 (TEMBLrel. 07, Created)
DT 01-AUG-1998 (TEMBLrel. 07, Last sequence update)
DT 01-MAR-2001 (TEMBLrel. 16, Last annotation update)
DE KIAA0517 PROTEIN (FRAGMENT).
GN KIAA0517.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RX MEDLINE=98290545; PubMed=9628581;

RA Nagase T., Ishikawa K., Miyajima N., Tanaka A., Kotani H., Nishio N.,
RA Ohara O.;
RT *Prediction of the coding sequences of unidentified human genes. IX.
RT The complete sequences of 100 new cDNA clones from brain which can
RT code for large proteins in vitro.*
RL DNA Res. 5:31-39(1998).
CC -!- SIMILARITY: CONTAINS A RING (C3HC4-CLASS) ZINC FINGER.
DR EMBL: AB011089; BAA25443.1; -
DR HSSP: P28990; ICHC.
DR InterPro: IPR000315; -
DR InterPro: IPR001258; -
DR InterPro: IPR001298; -
DR InterPro: IPR001841; -
DR InterPro: IPR002991; -
DR InterPro: IPR003649; -
DR Pfam: PF00097; ZF-C3HC4; 1.
DR Pfam: PF00630; Filamin; 1.
DR Pfam: PF00643; ZF-B_box; 1.
DR Pfam: PF01436; NHL; 6.
DR PROSITE: PS00518; ZINC_FINGER_C3HC4; 1.
DR SMART: SM00502; BRC; 1.
KW Zinc-finger.
FT NON_TER
SQ SEQUENCE 792 AA; 86659 MW; DA01F403BE580X08 CRC64;

Query Match 5.1%; Score 127.5; DB 4; Length 792.
Best Local Similarity 22.1%; Pred. No. 0.081;
Matches 89; Conservative 43; Mismatches 133; Indels 137; Gaps 18;

QY 66 TETELRI--KEIEKL-----DOQIPRP-----FTHYNTTSA 97
DB 314 TETELVLYKKQMSKLNELADQDFLHPRENDQDIFVETBGLKSHINLCTILTINAVA 473
QY 98 THSTATILNPDIYCGDQHLILEVRDHLRRKQYGGDFLRARMSPALMAGASKVTD 157
DB 374 SETVATGEGLEKQITII-GQPMSTVITTKDKDELCKTGNAYLTAEISTPD-GSVADEGIID 431
QY 158 FNNGTIVLSFTLFMEQVSLSL-----IHP-SEGVSLMGNARND 197
DB 432 NKGTYEELTYQKEDFTLSRLYDQIHRSFPFKLYIRSDVSPFTTEGKKRVKSPGS 491
QY 198 GY-----DRVIF-----TGQFVNGTSVHSECG--LIL 223
DB 492 GHVKQKAVKRPASMYSTGKRKENPIEDDLIFRVGTGRNKGFTNLQGVAASTSGILIA 561
QY 224 NTNAELCYLDNRDQ-EGFYCVRPQ-----HMPCAALTMNS-----KNKAVSYLSNQ 270
DB 552 DSNNGCVQIFSDNGQFKSRFGIRGRSPQQLDRPTGYAVHPSDIIILADYDNKWSIFSSD 611
QY 271 EKSLFERNVGVETMEKENTISVSKNTLKSVDLHESGKLQHLAVDLDRN--INIQWCK 328
DB 612 GKF---KTKIG-----SGKLMPKGVSVDRNGHIIYVDRK 644
QY 329 YCYPLIGSMYSVKEMEYLTRAIDRTGGEKNTVIYISLQGF 370
DB 644 ACCVFIFQPNKGI-----VTRFGSRNGNDQFA-----GPHF 675
PRT: 3603 AA.
ID P94459 PRELIMINARY; PRT: 3603 AA.
AC P94459:
DT 01-MAY-1997 (TEMBLrel. 03, Created)
DT 01-MAY-1997 (TEMBLrel. 03, Last sequence update)
DT 01-MAR-2001 (TEMBLrel. 16, Last annotation update)
DE PEPTIDE SYNTHETASE ORF4.
GN P94 OR PPSD.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/staphylococcus group; Bacillus.
OX NCBI_Taxid=1423;
RN [1]

Query Match	4.58;	Score 112.5;	DB 14;	Length 4060;
Best Local Similarity	19.68;	Pred. No. 14;		
Matches 103; Conservative	85;	Mismatches 213;	Indels 125;	

OY	22	IIYFQNSKRWASALNLSISLHAWNSKSLFPPKPLISAKLETETLRKEILEKLD-80	
Db	1614	IIETDLQKST--FTALYHPIKTH--NPKYAPQPOCKTLKMPPTQYISDDLOLEL-1668	
OY	81	-----COIPRPPTIVNIT--SAHSATILNPR-----108	
Db	1669	GPKHINFEQGVDTNRNTTLHANVTSQDIVFTTGTGEYVDSNHCQNAVITKRTTN-1720	
OY	109	-DYCRDQDHLILEYRDLERR-KYGGDFLRAMSSPALAGASGYTFNNGYLVS-166	
Db	1729	STTYILFSDTLTLAQRNKSQSHMHNTGHRFYNAELIPQTNKG-HMIPSLGQAVYA-1786	
OY	167	FTLMEQVSLSLILHPSEGVASLANSNOGY--GRVIFTQ--FVNGTSQVH--216	
Db	1787	NAELOTTOIAMLD-DTYEQAINAVYAKRIVSYYPNDIIGDASPMWCYINPSTSPSL-1845	
OY	217	-----SEGLINHTAEQCYLDND-----QSEFYCRQHMPCALHTMYSKKKVS-265	
Db	1846	ALPYSVEATFLTHGNI--TYTHDHYTHTEEFKALISLAPGYSTALSNGE--1901	
OY	266	YLSKQKSLFERSNVGVEIMEKFTSVS---KNTLKSVDLHESGKIQHQLAVLDLBN-321	
Db	1902	--TYARNPNPDQGTGKTLHPRIINTHTIMELATINILKTA-ITAVAKIRPALATL--G-1956	
OY	322	INIGMOKYCYPL-----GSMNTYSVEMEXYLRALDTSGEKNTYIVISLQHRREPPD-376	
Db	1957	LSPDFINATHPRLHINTYLGSL-ITHHKLATLPLAKCA---SGVAVNFCVN---TPIN-2008	
OY	377	VFIRRALNTHKAIQHLILSPDINVIKTE-----NIREM--YNDAPRSDFHGY-424	
Db	2009	CTSDKGYGINSQSLSELEHNTQNTAINTELPYTLGSDVRDMCLYNSIRLADSNQY-2068	
OY	425	IQYLINDIPQDLSIIDADIVTYGNNVH-----PPQH-461	
Db	2069	-----STSPFDADITDITPFR-KNSVPSRICYSKIPPIH-2100	
RESULT	6		
09YT27		PRELIMINARY:	PRT: 688 AA.
ID	09YT27		
DT	01-MAY-2000	(TREMBLrel. 13, Created)	
DT	01-MAY-2000	(TREMBLrel. 13, Last sequence update)	
DE	01-OCT-2000	(TREMBLrel. 15, Last annotation update)	
DE	F58E2.3	PROTEIN.	
GN	F58E2.3.		
OC	Caenorhabditis elegans.		
OC	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoides;		
OC	Rhabditidae; Peloderinae; Caenorhabditis.		
OX	NCBI_TaxID=6239;		
RA	SEQUENCE FROM N.A.		
RA	STRAIN=BRISTOL N2;		
RA	MEDLINE=94150718;		
RA	Wilson R., Almscough R., Anderson K., Baynes C., Berks M.,		
RA	Bonfield J., Burton J., Connell M., Cosey T., Cooper J., Coulson A.,		
RA	Crafton M., Dear S., Du Z., Dublin R., Favetto A., Fulton L.,		
RA	Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,		
RA	Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,		
RA	Lightfoot J., Lloyd C., Mowbray A., Mortimore B., O'Callaghan M.,		
RA	Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkneen R.,		
RA	Smalton N., Smith A., Sonhammer E., Staden R., Sulston J.,		
RA	Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,		
RA	Watson A., Wellstock L., Wilkinson-Sprat J., Wohlman P.,		
RT	*2.2 Mb of contiguous nucleotide sequence from chromosome III of C.		
RT	elegans.		
LN	Nature 368:32-38(1994).		
LN	[2]		

RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Goela D., Delehanthy A.;
RT "The sequence of *C. elegans* cosmid F5862.",
RL Submitted (OCT-1998) to the EMBL/GenBank/DBS databases
11

Query Match	4.48;	Score 109;	DB 5;	Length 688;
Best Local Similarity	19.38;	Pred. NO. 2.1;		
Matches 112;	Conservative 76;	Mismatches 200;	Indels 192;	Gaps 27

QY	1	MSNMANKSLALLLETLASMIIFIVONSTKWSALNLSLHTR-----NNSIKSLF	64
Dd	93	LEISLDLDEL-----DKRMSASOSIOFYDMVIGLPLNNHVGOLK	114
QY	55	PKPIPLSL-----KPLTELEIKREIKELKQVIVR	84
Dd	134	PK--LISLGFQVQALSLFHHISSNHEILINIFRKEFTELEHHDEIYKM--KQKHA	144
QY	87	PEYVNTTTSNTHSTATT--LNPRTYCRGGQIHLILEYRCHLIRKKVYOGIPLAARNS	144
Dd	190	KTLHLEECSTSIENVLHFGAGAFKVDVISTDIDQLMKELSTHSAITLKFNISF-----	244
QY	145	PALMAGASGVTD-----FNGGTVLSPFLFW--EGQVSLILLIHSEGV--SALMS	194
Dd	243	-----NNADSTIDESLTLGLFQSPKKHSLFMESIDGRIECSSEFLTKSTIYCPKACEE	296
QY	194	AARQGYDVIFITQGFNGTSGVHSEC-----GLILNT-----NAELQYTLNRQEG	244
Dd	299	LNR-----VSSKELECTSLCTSELSIKYHLNGHLMNMLSHGSDIDCKL--PRISM	354
QY	241	FYCVAPHQMCALTHMYSKNKKYSYLSQKSETERSNVGEVMEKENTISVSK-----	294
Dd	351	V-----QKCVOLI-----KPAELIALEINLINVENSVSI--FINAKIATHE	394
QY	297	-----NTLKSVDLHESGLOHQLAVDLDRNINIGW-----KSYVPLISNM	344
Dd	395	IDLTTSYODHYSKVIOKVEKIRHNGLVRELC-----SVNRMQIGFSLSKFADLKLMI	444
QY	339	VSQKEME-----YLRATRIDTGENKTVVI-----SLGSHFRPTIIV	384
Dd	450	AQKSEIDQVRLSDPLSIKANSPGSKNKHVLPRFLAVALLDVIYSTIQQDLNGLPYKLS	504
QY	380	KALNVAKRAIOLHLLSPDTPVVIKTEINREKYNDAERFSDFHOYIVLCLIKPLFLSV	434
Dd	510	IQVNLQETLQIOLLSLRPS--LKTEIEIKLKD-----ILSKSLKLNEA	553
QY	440	SIIDAM-----DITIAGTINNVHPO-----HVGQNIIL	470
Dd	554	AQDDQMKNAEDILIELT---APQGIQYFLPHVNEINIM	590
RESULT	7		
ID	09G276	PRELIMINARY;	PRT: 1387 AA.
AC	09G276;		
DT	01-MAR-2001 (TREMBLrel. 16, Created)		
DT	01-MAR-2001 (TREMBLrel. 16, Last sequence update)		
DT	01-MAR-2001 (TREMBLrel. 16, Last annotation update)		
DE	RETICULOCYTE-BINDING PROTEIN 2 HOMOLOG A (Fragmen?)		
OS	Plasmodium falciparum.		
ES	Eukaryota: Alveolata: Apicomplexa: Hemosporida: Plasmodium.		

OX NCBI_TaxID=5833;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20402589; PubMed=10920203;
 RA Hayer J.C., Galinski M.R., Ingravallo P., Barnwell J.W.;
 RT "Two plasmodium falciparum genes express merozoite proteins that are
 related to Plasmodium vivax and plasmodium yoelli adhesive proteins
 involved in host cell selection and invasion."
 RL Proc. Natl. Acad. Sci. U.S.A. 97:9648-9653(2000).
 DR EMBL: AF196347; AAF98066.1;
 FT NON_TER 1
 SO SEQUENCE 1387 AA; 164871 MW; 5660154744F0C285 CRC64;

Query Match 4.4%; Score 108.5; DB 5; Length 1387;
 Best Local Similarity 20.2%; Pred. No. 6.4;
 Matches 80; Conservative 70; Mismatches 137; Indels 109; Gaps 19;

QY 140 ARMSPALMAGAGKYTDEN--NGTIVSTFLTEGQVSLSLIHPS--GVSAIWSARN 196
 DB 250 ANFGSIIWMSDITETITPENPLENDLNLQITFEKKEITSTLENSDLELDHIGSND 349
 QY 197 GGYDRAVITGQFVNSTQVHSEGLINTNAELQYLDRPQEGFVCRQHPQCAL-- 254
 DB 350 ESIINL---KYNDILIELHT---YSTULKYLDNIQKLGDC--NDLVKCKELRE 397
 QY 255 --THMSKNKKVSYLSKQKSL-----FERSNVGEIWEKFNPTIS 292
 DB 398 LSTALYDLKIOITSVINKEINDISNNIDVSNKLINEIDAIQVNFKEIKELFDVNEEKTL 457
 QY 293 -----VSKCNTLSVDLHESGKLOHQLAV---DLDRNINIMOKVCYPLIGSMYSVK 342
 DB 458 DTKNATIVKKAELKNVDIN---KTKEDDIYFNLDL-----LEKSLTSSN 502
 QY 343 EMEYLT-----RAIDRTGKKNIVYISLGOHFRPPIDVFRALVNHKAI 389
 DB 503 EMEKTIWONSYSFSDINKNINDIKEMKTLIP-----LDLINSNGHIDISL 552
 QY 390 QHLLRSPTVITIKTENIRKMYNDARESDPHGTLQ--YIIKDFQULSVSIIDADI 447
 DB 554 YNFITKNIQIKIGNDIKIRKENDINICEF---YIDNNFNFKS-----DISIFNKYDD 604
 QY 448 TIA---YGTNNVHP-----POHVVGNQINILN 472
 DB 605 HIKVDNTYSNNIDVNNKNSLSEHVI--NATNIEN 639

RESULT 8
 ID 087704 PRELIMINARY; PRT: 2571 AA.
 AC 087704;
 DT 01-NOV-1998 (TREMBlrel. 08, Created)
 DT 01-NOV-1998 (TREMBlrel. 08, last sequence update)
 DT 01-MAR-2001 (TREMBlrel. 16, last annotation update)
 DE FENGYCIN SYNTHETASE.
 GN FEND.
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Bacillus.
 OX NCBI_TaxID=1423;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=F29-3;
 RA Huang S., Chen C., Lui S.;
 RT "Analysis of fengycin synthetase gene fend.*;
 RT Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL: A101149; CAA09819.1;
 DR HSSP: P14687; JAM0;
 DR ITCRPT0: IPR000255;
 DR InterPro: IPR000873;
 DR InterPro: IPR001242;
 DR Pfam: PF00501; AMP-binding: 2.

DR Pfam: PF00550; pp-binding: 2.
 DR Pfam: PF00668; Condensation: 3.
 DR PRINTS: PR00154; AMPBINDING.
 DR PROSITE: PS00075; ACP_DOMAIN: 2.
 DR PROSITE: PS00455; AMP_BINDING: 2.
 DR PROSITE: PS00012; PHOSPHOPANTETHEINE; UNKNOWN_1.
 KW Phosphopantetheine.
 SO SEQUENCE 2571 AA; 290799 MW; 83098B3ADAABD48 CRC64;

Query Match 4.3%; Score 106; DB 2; Length 2571;
 Best Local Similarity 19.3%; Pred. No. 25;
 Matches 95; Conservative 71; Mismatches 172; Indels 154; Gaps 25;

QY 109 DTTCGQQLHL-LEVRD-----HIGRR-----KQYGDFLRARMSPAL-- 147
 DB 1199 DLVC-GKELPALNLHKDFAVWQCKEPPEHYSKQAVWLMQLG-----SLVYLDLP 1250
 QY 148 MAGAGKVTDFNNGTLYSF-----TLFEGQVSLSLIHSEGVSAIWSARNOG 198
 DB 1251 LDKTRPLLPFSGCTFEVNIIDKRTAEHLRHLMEFTGTMYMLL---AVYSIMLSKLSG 1306
 QY 199 YDRAVITGQFVNSTQVHSE--CGILINTNAELQ-----YIDNRDQGFYCVAPQ 247
 DB 1307 QEDIV-VGSPAAGRPHADLERVIGMEFNTLAMKQCPQCKRTFSSYQIETRELATAYEHQ 1365
 QY 248 HMPQALTHWYSKNKKV-----SYLSKQKSLFERSNVGEIWEKFNPTISVSKCNTLK 300
 DB 1366 DYPEELVNLKREKREVRNPLFPALMAYIDNSDQFQVPELSTSYTSVSHVNSFDLTL 1425
 QY 301 SVDLHESG---KLOHQLAVDLDNRNINIMOKVCYPLIGSMYSV---KEMEYLTFRADR 353
 DB 1426 HAEHSNGIRCREYSTALFEETI--ARASHFELYVIGTISDIHKKLSEMOILS----- 1479
 QY 354 TGSKRNTVYISLGOHFRPPIDVFRALN-----YHKAI--QHLL----- 394
 DB 1480 --APARELLETMGQ--YADYRDESTIVRLFQKAHEHRTAVVCHILTYRLINKAE 1536
 QY 395 -----RSPDTM-----IKT-----ENIDEMTADA 415
 DB 1537 RTATMLMKQVSGIIGLMDRSPMVTYGVLSIKKAGVYLDIDEPYQERISPMNDIS 1596
 QY 416 -----ERSDFHYOYLIIKDFQULSVSIIDADIT-----IAYGTNNVHP 458
 DB 1597 GARILLTEGHKKRPADYHGOILYL--NDAENELISPLKQETLADQPAVYIYTSCTTG 1654
 QY 459 POHVVGNQINIL 470
 DB 1655 PKGVVVEHNRVY 1666

RESULT 9
 ID 013788 PRELIMINARY; PRT: 3262 AA.
 AC 013788;
 DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, last sequence update)
 DT 01-JUN-2000 (TREMBlrel. 14, last annotation update)
 DE APOIIPROTEIN B-100 (FRAGMENT).
 GN APOB.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=87191999; PubMed=2883086;
 RA Carissom P., Darinfor C., Olotsson S.O., Bjurzell G.;
 RT "Analysis of the human apolipoprotein B gene: complete structure of
 the B-74 region.";
 RL Gene 49:29-51(1986).
 DR EMBL: M15421; AA51758.1;
 FT NON_TER 1

SEQUENCE 3262 AA: 370140 MM: 56603BC0618DD40D CRC64:

Query Match 4.2%; Score 104.5; DB 4; Length 728;
Best Local Similarity 20.5%; Pred. No. 46;
Matches 89; Conservative 72; Mismatches 169; Indels 105; Gaps 22;

QY 56 KTPILSLKPLTEFE-LRIKEIIEKLDQOIPRPPTHV---NTTSATHS-TATILNPRD 109
DB 744 KVPILSEPINIDALEMRDAVEK-----POEFTIVAFVKDKQDVHSINLPFETIQ 797
QY 110 TYCRGDQHLILEVRDHLGRKQYGD-FLPARSSPALMAGASGVDPFNGGYLVST 168
DB 798 EYERNROTIIIVLENVQRNLKHINIDQFVRK-----YRAALCKLHQAN-DYLNSTN 849
QY 169 LFMEOVSLSLILHPSSEVSALMS-----AANQGYDRI 204
DB 850 --WEROVS-----HAKKLTALTLYKRYITENDIQALDANKINENKLSQLOTY--MIQ 899
QY 205 TGOEYNGTSOVHSECGILINTNAELCOYLDRQDEGFYCVPRQHPICALTHMYSK---- 260
DB 900 FDOYIKDSYDLHDKIALIANIIDELIEKLSLDEH--YHIRVNLVKTIHDLHLEFINIDF 957
QY 261 NKKVSYLSKQKSLFERSNVGEIEMKNTISVSKNTLSYDL-HESGKL-QHOLAVDL 318
DB 958 NKSSTASWIONVDTKYQIRIOIEKLODK---RHIONIDQHLAKLQKHEIALIV 1013
QY 319 DNINIQOKYCPPLIGSMTYSVEMELTRADRTGKGNVVISLQHFPPPI--- 375
DB 1014 -----RYLLDQGLT-----TISFERI---NDVL-----EYVNHVYINLI 1044
QY 376 -DVFIRALNVHKAIOHLLLSRPT-----MYIKTEINREMYDARFSDFGIYVLI 429
DB 1045 GDFEYAEKINAFRAKVKHELIEREYDQIOYLMKLVLAHQYKLTQKLSNVLQGVK 1104
QY 430 IKDIFQDLSYSIDA 444
DB 1105 IKDYFEKLVGFTIDA 1119

RESULT 10
ID 013779 PRELIMINARY: PRT: 728 AA.
AC 013779;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)
DE APOLIPROTEIN B-48 (FRAGMENT).
GN APOB.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_Taxid-9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-88050832; PubMed-3676265;
RA Hardman D.A., Procter A.A., Chen G.C., Schilling J.W., Sato K.Y.,
RA Lau K., Yamanaoka M., Mikita T., Miller J., Crisp T.M., McEnroe G.,
RA Scarborough R.M., Kane J.P.;
RT "Structural comparison of human apolipoproteins B-48 and B-100.";
RL Biochemistry 26:5478-5486(1987).
DR EMBL: M17367; AAA51741.1; -;
FT NON_TER 1
FT TER 728
FT NON_TER 728
SQ SEQUENCE 728 AA: 83217 MM: CF27745AEA02B699 CRC64;

Query Match 4.2%; Score 104; DB 4; Length 728;
Best Local Similarity 20.5%; Pred. No. 5.8;
Matches 89; Conservative 72; Mismatches 167; Indels 106; Gaps 23;

QY 56 KTPILSLKPLTEFE-LRIKEIIEKLDQOIPRPPTHV---NTTSATHS-TATILNPRD 109
DB 744 KVPILSEPINIDALEMRDAVEK-----POEFTIVAFVKDKQDVHSINLPFETIQ 797

DB 355 KVPILSEPINIDALEMRDAVEK-----POEFTIVAFVKDKQDVHSINLPFETIQ 408
QY 110 TYCRGDQHLILEVRDHLGRKQYGD-FLPARSSPALMAGASGVDPFNGGYLVST 168
DB 409 EYERNROTIIIVLENVQRNLKHINIDQFVRK-----YRAALCKLHQAN-DYLNSTN 460
QY 169 LFMEOVSLSLILHPSSEVSALMS-----AANQGYDRI 204
DB 461 --WEROVS-----HAKKLTALTLYKRYITENDIQALDANKINENKLSQLOTY--MIQ 513
QY 205 TGOEYNGTSOVHSECGILINTNAELCOYLDRQDEGFYCVPRQHPICALTHMYSK---- 260
DB 511 FDOYIKDSYDLHDKIALIANIIDELIEKLSLDEH--YHIRVNLVKTIHDLHLEFINIDF 957
QY 261 NKKVSYLSKQKSLFERSNVGEIEMKNTISVSKNTLSYDL-HESGKL-QHOLAVDL 318
DB 569 NKSSTASWIONVDTKYQIRIOIEKLODK---RHIONIDQHLAKLQKHEIALIV 624
QY 319 DNINIQOKYCPPLIGSMTYSVEMELTRADRTGKGNVVISLQHFPPPI--- 375
DB 625 -----RYLLDQGLT-----TISFERI---NDVL-----EYVNHVYINPY 655
QY 376 -DVFIRALNVHKAIOHLLLSRPT-----MYIKTEINREMYDARFSDFGIYVLI 429
DB 656 WDFEYAEKINAFRAKVKHELIEREYDQIOYLMKLVLAHQYKLTQKLSNVLQGVK 715
QY 430 IKDIFQDLSYSIDA 443
DB 716 IKDYFEKLVGFTIDA 728

RESULT 11
ID 043988 PRELIMINARY: PRT: 800 AA.
AC 043988;
DT 01-JUN-1998 (TREMBLrel. 06, Created)
DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE HOMEOBOX-CONTAINING PROTEIN WARIAL (FRAGMENT).
GN WARIAL.
OS Dictyostellium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyostelidia; Dictyostelium
OX NCBI_Taxid-44689;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-KAX3;
RA Han Z., Firtel R.A.;
RL Development 0:0-0(1998).
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -1- SIMILARITY: TO OTHER HOMEOBOX DOMAINS.
DR EMBL: AF036170; AAB92245.1; -;
DR HSSP: P42711; 1917.
DR InterPro: IPR001356; -;
DR InterPro: IPR002110; -;
DR Pfam: PF000023; ank; 9.
DR Pfam: PF00046; homeobox; 1.
DR PROSITE: PS50088; ANK_REPEAT; 8.
DR PROSITE: PS50297; ANK_REPEAT_REGION; 1.
DR PROSITE: PS00027; HOMEOBOX_1; 1.
DR PROSITE: PS50071; HOMEOBOX_2; 1.
DR SMART: SM00389; HOX; 1.
KW DNA-binding; Homeobox; Nuclear protein.
FT NON_TER 800
FT TER 800
SQ SEQUENCE 800 AA: 86723 MM: 7FACC2C09884DB4F CRC64;

Query Match 4.2%; Score 104; DB 5; Length 800;
Best Local Similarity 18.4%; Pred. No. 6.7;
Matches 83; Conservative 68; Mismatches 136; Indels 144; Gaps 14;

QY 104 IINPDP-----TYCRDQHLILEVRDHLGRKQYGD-FLPARSSPALM-AGASGVDP 157
DB 343 LHPNIDFLIRINPDPDPSIDVNAKSGKSLFLTAALVLEYOVRLIESANINIKD 402

```

OY 158 FNNQIVLSTFLFMEGQVSLILLIHS-----EGVSAIWSARNOC-----YD 200
1b 403 NQGDIVLIASVL-GNQP IVELLEHRAEDPNLVNDGVSPLFSACKGHQIXSLIIDD 461
OY 201 RVIFTGQFVNGTSQVH-----SECGILLNTNALCYLDNRDQEGFYCAPRQMPCA 252
Db 462 REVSVKTKINGETPLHIAISKGFECIKKILLIETEAQ-ASVIDSNRR-----TPLHAC- 513
OY 253 ALTHMYKNNKVSYSLSKOKSLIFSVNGVEIMENKNTISVSKCNTLKSVDLHESGKLOH 312
Db 514 -----IMGYSTIAKLJCN-----GADMN-----532
OY 313 QLVADLRNINIQMOKYCYPLIGSMYSVEMEYLTRAIDRTGCKNTVYISLGGHPP 372
Db 533 --AIDIDGHTPLH-----TSSLGQCYLITRILLEAGDPN-----IODSGCTP 574
OY 373 PPIDVFIIRALNVHKAIOHLLRSP-----DTMVITKENTREMYDAERF-----418
Db 575 -----IHVAVRESRIETVKFLIFNSKILNTKNGQNLHLVSQFASLMMGQM 622
OY 419 -----SDFHSY-----IOYLIRKD-----IFODLSYSI 441
Db 623 IFESKGEIADSDSDGQTPYLAAGAKTNFVKYLLSKGRSKKIRPLEKLIQENQDMEI 682
OY 442 IDAMDITAYGTNNHPPQHVVNGQINILLN 472
Db 683 IOMLESTVTKSSNNNSNSNT--NNNNNN 711

RESULT 12
OYVS73
ID OYVS73 PRELIMINARY: PRT: 602 AA.
AC OYVS73
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE G08546 PROTEIN.
GN G08546.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanalides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Morlan J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Ajay J.F., Abmayyan A., An H.-J., Andrews-Pfankuch C., Baldwin D.,
RA Baller R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borokova D., Botchan M.R., Bouck J., Brockstein P., Brothier P.,
RA Burlis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos H., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Lodon K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Petraz C., Ferlita S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Gang N.S., Gelbart W.M., Glasser K.,
RA Glodok A., Gong F., Gottell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibeagwan C.,
RA Jajuli M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko F., Lai Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei H., McIntosh T.C., McLeod M.P., McPherson R.,
RA Morkulov G., Mitsuina N.V., Modarri C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,

```

```

RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacheb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Peltan J., Puri V., Reese M.G.,
RA Relbert K., Remington K., Saunders R.D.C., Schaefer F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spler E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svitskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao O., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL: AE003559; AAF50551;
DR Flybase: FBgn0035785; G08546.
DR InterPro: IPR001873; -.
DR Pfam: PF00858; ASC: 2.
DR PRINTS: PRO1078; AMINACHANNEL.
DR SEQUENCE 602 AA; 68399 MW; 7CDA34CC9D78A22A CRC64;

Query Match 4.18; Score 102; DB 5; Length 602;
Best Local Similarity 20.98; Pred. No. 6.4;
Matches 86; Conservative 61; Mismatches 146; Indels 118; Gaps 23;

OY 9 KSLALLFIASWI-----IFTVQNSTKWSALNLSISLHYNN-----STKS 52
Db 122 RPLLERLF-----WISVLVSFTCVKLTILIMDK-----MNNPIVFAFKRPV 168
OY 53 LEFPTPLISKLTELEIRIKEIEIKDOOIPPRPF-THVTTTSATH-----STATI 104
Db 169 WQIPFPAVIVCPETKTR--REIFNFTDSHYOVRPOSVANGVYDSDRGMEITDALTE 225
OY 105 LNPR--DTYGRGDQILILEVRDLGRKQYGDPLARMSPPALMAAGKGVTDPPNNGT 162
Db 226 VSPHFDQTYLNCMKWRSPVKCSDFHFHFTVTEGCVFSPNSLPAEIRFREGIIPDF----281
OY 163 YLVSFTLF-MEGQVSLILLIHPSEGVSLMSARNQG-DVYIFGQ-----FNGCS- 213
Db 282 -----IFRENRLSMDW-----NVEDGYSA--SADISPPNNVLPQARRAGLYLFNGAEI 330
OY 214 QVHSEC-----GLINTNALCYLDNR-----DQGFYCVRPOHMCALTHMSKN 261
Db 331 DFDDMCRGPQVGFKILLTPGDVAOVSKQYFRIPPDQVILSIRPKITITSDGLKHEPN 390
OY 262 KVSYSLSKQEK-----SUPERANGVEIMENKNTISVSKCNTLKSVDLHESGKLOHOLAVD 317
Db 391 RRCYFQKEKELRYNINISQSNCELELANF---TLIKGCVK-----FS 432
OY 318 LDRNINIQMOKYCYPLIGSMY--YSVEMEYLTR-----AIDRTGCE 357
Db 433 MPRNVN-----PVCGDASIKCYQADDELLREFTOGLVNAAGENTGE 476

RESULT 13
ID 095151 PRELIMINARY: PRT: 1507 AA.
AC 095151
DT 01-FEB-1997 (TREMBLrel. 02, Created)
DT 01-FEB-1997 (TREMBLrel. 02, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE BRCA2 (FRAGMENT).
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RA Bignell G.R., Micklem G., Stratton M.R., Ashworth A., Wooster R.,
RA submitted (JUL-1996) to the EMBL/Genbank/DBJ databases.
DR EMBL: Z75564; CA99994.1; -.
DR InterPro: IPR002093; -.
DR Pfam: PF00634; BRCA2_repeat; 7.

```


DR PROSITE; PS50138; BRCA2_REPEAT; 6.
 FT NON_TER 1 1507 1507
 SEQUENCE 1507 AA; 168659 MW; 2C78ABA31B93E257 CRC64;

Query Match 4.0%; Score 99.5; DB 6; Length 1507;
 Best Local Similarity 20.6%; Pred. No. 39;
 Matches 77; Conservative 55; Mismatches 143; Indels 99; Gaps 18;

QY 22 IIFTEFONSTKYWSALNLSLHYMNSTKSLFKPTPLSLK-----PLTEEL-70
 DB 436 IISGFVONSTYV-----SDSESGHTAPPLLSLKQDFDSNRNLTPEQKAEIT 481
 QY 71 RIKEIIEKLDQOIPRP-----THV--NTTSAITSTAITIIPDTCRCDOHLHLLVYR 124
 DB 482 ELSTIIIESSGOFETQFRKPSHIIOKNFEPENOLTIINSTSKWKODDLHTTAPS 541
 QY 125 -DHGRRKO---YGGDFLRAMSSPALMAGAGKVTDFNNGTYLVSFTLLFNEGQ-VLSLT 179
 DB 542 ISQVDSKKSEGIIGGKKFACLSRTSCNHSASGYSTDKN-----VEFRGFYSAGTKLVN 597
 QY 180 LLIHPSEGYSALMSAR-----NOGYDRVIFTCGFVNGTSQY-----PLTEEL-215
 DB 598 -----GSEALQKRAKKLFSDLENIMEFISYVDNRSSSKYNDYSMTQLEDONKNLN 650
 QY 216 --HSECGLLTNALC-----QYLDNRQDSFYCVRPQHPCAALT--HMYSKN 261
 DB 651 EPNKRCRLILONNILEMTDIFVEEYTESYRNTENEGNCTDAGRNTCSNDSGDSKKN 710
 QY 262 KKVSYLSKQKSL--FERSNVGEIEMKF---NTISVSKNTLSKVLHESGKLOH---312
 DB 711 DTV-YIHEPENGCPIDQHNHIDLKLFSSQFMKEGNTQIKESLSTLCLEVMKAEITSHVM 769
 QY 313 ----OLAVDLRNI 322
 DB 770 SNKQOLLANTGONI 783

RESULT 14
 ID 096167 PRELIMINARY; PRT; 795 AA.
 AC 096167;
 DT 01-MAY-2000 (Tremblrel. 13, Created)
 DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
 DE 01-MAY-2000 (Tremblrel. 13, Last annotation update)
 DE HYPOTHETICAL 90.2 KDA PROTEIN.
 GN T3J7.30.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
 OC Brassicales; Brassicaceae; Arabidopsids.
 OC NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Choisme M., Robert C., Bottler P., Wincker P., Cattolico L.,
 RA Artiguenave F., Saurin W., Weissenbach J., Mewes H.W., Mayer K.F.X.,
 RA Lemcke K., Schueller C., Queller F., Salenoubat M.;
 RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Arabidopsis sequencing project:
 RA Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: A1049746; CAB41853.1;
 KW Hypothetical protein.
 SO SEQUENCE 795 AA; 90232 MW; 3E4A3EBD36105088 CRC64;

Query Match 4.0%; Score 98.5; DB 10; Length 795;
 Best Local Similarity 19.1%; Pred. No. 18;
 Matches 100; Conservative 79; Mismatches 181; Indels 163; Gaps 26;

QY 33 VMSALNLSLHYMNSTKSLFKPTPLSLKPLTE-----LRKEIIEKLDQO 82

DB 149 VMSLKNL-----RTSRSGFEFVLAHNTLTTELLISVAKHRRMAKLSAVNR 20
 QY 83 I-----PPRFTVNTTSAT--HSTATIINFRDYVNR 119
 DB 204 VDRALAMMPQADYRALLSLRMPQ-----LSTLSASLSKSNVQNR-----259
 QY 117 LHILLEVRDHLGRKKQYGGDFLRAMSSPALMAGAGKVTDFNNGTYLVSFTLLFNEGQ-VLSLT 179
 DB 251 ---LTFMEGSL--KSYCGSF--HALCSLOGIOLQRKRROGRIKGBENLPHOTMAHEL 404
 QY 177 LSTLLHPSREGYSALMSARNGYDVIF--TQYVNGTSVISECLINNAHLYLT 234
 DB 305 VNPFLV-ASQHPFTKWEKPEFIFALVYKTRIDVDSMDL-----LPIVY 480
 QY 235 NNDGEFYCVRPQHPCAALTHV-----SKKKVSYV-----487
 DB 351 EARLGYSTCREWYSAVSSLSLYLVKEIFPIYVQGLDENETDLRSFAKSKHILDM 410
 QY 268 ---SKQESLFRSNVGEIEMKFNTISVK---CNTLSVYLHESGKLAHQA-----415
 DB 411 ISDFKVSLSVSGI--LSLQEDGNLRTSLSYECRDMDLMAEILEDEKLVKKE 419
 QY 316 VLDKNININQWK-----YCPPLISMTYSKEMEYLTATIRISKEKNIYVLSL 437
 DB 470 IINDKNMTAKVQDELISSNVYRPPLISIF-----LQHLSTIERS-----KSVFALYLR 500
 QY 368 QHRPEPIDVFIRBALN--VHKAIQHLLRSFDT--MYIITKEN;KEMNMAHRSYHRY 434
 DB 521 AR-----FLRLASPTIHKFLDCLLRDODDGLTALTEN-----NCLIKVSN INR 477
 QY 425 IOTLIIRKIFODLSVITIDAMDITAY---GTNNVHPQHYVQ 464
 DB 568 GHY-----ESYLEMSEDDVFLEMGTQV-HDQPEVQV 539

RESULT 15
 ID 096167 PRELIMINARY; PRT; 1004 AA.
 AC 096167;
 DT 01-MAY-1999 (Tremblrel. 10, Created)
 DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)
 DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
 DE SERA ANTIGEN/PAPAIN-LIKE PROTEASE WITH ACTIVE SER.
 GN PFB0360C.
 OS Plasmodium falciparum.
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OC NCBI_TaxID=5833;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE:99021743; PubMed:9804551;
 RA Gardner M.J., Jettellin H., Garucci D.J., Cummings L.M., Atavid L.,
 RA Koehn E.V., Shallow S., Mason T., Yu K., Fujii G., Petersen J.,
 RA Shen K., Jing J., Aston C., Lai Z., Schwartz D.C., Pertea M.,
 RA Salzberg S., Zhou L., Sutton G.G., Clayton R., White O., Smith H.O.,
 RA Fraser G.M., Adams M.D., Venter J.C., Hoffman S.L.;
 RT "Chromosome 2 sequence of the human malaria parasite P. falciparum".
 RL Science 282:1126-1133(1998).
 DR EMBL: AE001389; AAC71861.1;
 DR InterPro: IPR000169;
 DR InterPro: IPR000668;
 DR Pfam: PF00112; Peptidase_C1; 1.
 DR PRINTS: PR00705; PAPAIN.
 DR PROSITE: PS00640; THIOL_PROTEASE_ASN; 1.
 KW Hydrolase; Protease; Thiol protease.
 SO SEQUENCE 1004 AA; 116573 MW; 1F1750BA2A9HE10C9 CRC64;

Query Match 4.0%; Score 98.5; DB 5; Length 1004;
 Best Local Similarity 17.5%; Pred. No. 26;
 Matches 85; Conservative 88; Mismatches 151; Indels 191; Gaps 42;

```

OY 23 IFTFONSTKWSALNLSISLHYMNSTKSLFPKPTPLISKPLTETELRIKETEKIDQ 82
Db 71 IHASTSOTKEMSPNNLT-----SKKKKKKEIRPKDIMSND- 109
OY 84 IPRPPTHVNTTISATHSTATILNPRDTCRGDOLHILLEVRDHLGRKOYGGDFLRAM 142
Db 110 -----SNTSSINKON-----NNQIKSVL-LKENKG-----VKI 136
OY 143 SSPALMAGASGYTDFNNNGTIVLSTFLFEGOVSLSLIT-HPSEGVSAIAMSARNOGYDR 201
Db 147 TGPC-----NNLSIFLPHIYIDVETRYNNIELKVEL 169
OY 202 VIFTG--QFVNGTSOYHSEGLINTNAELCOYLDRDQEGFYCVRPOHPCALTHMYS 259
Db 170 DEFSUSIKFKDITTELRTSDOTLMTNPNVGSRDITLQDRLYN-----CA----- 216
OY 219 KKKKVSYSKQEKSLFERSNNGVEIMEKFNITISVSKCNTLKSVDLHESGKLOHOLAVDLD 319
Db 217 ENKTFKFEV-----VIKDN-----LTLK-WKVEYETGVYNNK--VDIR 251
OY 420 RNINIQKQYCPPLIGSMYSYKEMEYLTPRAIDRTGGEKNTVIVISLQHFREFPIDVFI 379
Db 252 Q--YKMKELTPTITTOIHSYSEKD-THLE---SKNYVIKTDIPETCDVMATNGEL 303
OY 360 RRALNVKAIQ-HILIRSPDT---MVIKTENIREMYN-----DAERPSDFH----- 422
Db 404 SGNINIEKLECTLLVONNDTSECEFTYVSNQVRENFNOIKAEAEDEDFRNYHLDITIN 363
OY 423 -----GYIOYLIIKDIPODLVSIIIDAMDITIAVGTNNVHPPOHYVGNQINI 469
Db 364 NILKRIYIKNKNGKKELITLELONFLKESITDYCKILREIDTNGT-LVNHELGNNDV 422
OY 470 LNTYI 474
Db 423 PNNLI 427

```

Search completed: October 12, 2001, 16:05:18
 Job time: 161 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM Protein - protein search, using sw model

Run on: October 12, 2001, 16:03:25; Search time 34.67 seconds

(without alignments)
956,485 Million cell updates/sec

Title: US-09-729-454-2

Perfect score: 2908

Sequence: 1 MMSNTMLQKTLILILISFSVV.....IHPDHVIGNQIMFLNYIC 547

Scoring table: BLOSUM62

Gapop 10.0, Capext 0.5

Searched: 412676 seqs, 60623988 residues

Total number of hits satisfying chosen parameters: 412676

Minimum hit seq length: 0

Maximum hit seq length: 2000000000

Post-processing: Minimum Match 0%

Listing list 45 summaries

Database:

A-Geneseq-0601:
1: /SID8/gcadata/geneseq/geneseq/AA1980.DAT:
2: /SID8/gcadata/geneseq/geneseq/AA1981.DAT:
3: /SID8/gcadata/geneseq/geneseq/AA1982.DAT:
4: /SID8/gcadata/geneseq/geneseq/AA1983.DAT:
5: /SID8/gcadata/geneseq/geneseq/AA1984.DAT:
6: /SID8/gcadata/geneseq/geneseq/AA1985.DAT:
7: /SID8/gcadata/geneseq/geneseq/AA1986.DAT:
8: /SID8/gcadata/geneseq/geneseq/AA1987.DAT:
9: /SID8/gcadata/geneseq/geneseq/AA1988.DAT:
10: /SID8/gcadata/geneseq/geneseq/AA1989.DAT:
11: /SID8/gcadata/geneseq/geneseq/AA1990.DAT:
12: /SID8/gcadata/geneseq/geneseq/AA1991.DAT:
13: /SID8/gcadata/geneseq/geneseq/AA1992.DAT:
14: /SID8/gcadata/geneseq/geneseq/AA1993.DAT:
15: /SID8/gcadata/geneseq/geneseq/AA1994.DAT:
16: /SID8/gcadata/geneseq/geneseq/AA1995.DAT:
17: /SID8/gcadata/geneseq/geneseq/AA1996.DAT:
18: /SID8/gcadata/geneseq/geneseq/AA1997.DAT:
19: /SID8/gcadata/geneseq/geneseq/AA1998.DAT:
20: /SID8/gcadata/geneseq/geneseq/AA1999.DAT:
21: /SID8/gcadata/geneseq/geneseq/AA2000.DAT:
22: /SID8/gcadata/geneseq/geneseq/AA2001.DAT:

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	103.5	3.6	1237	21	AAV6999	S. cerevisiae esse
2	101	3.5	452	21	AAV70502	Corn tariesyllitans
3	97	3.3	252	21	AAAB01392	Neuron-associated
4	97	3.3	252	21	AAV99401	Human PRO1327 (UNQ
5	97	3.3	252	22	AAAB7571	Human PRO1327. Ho
6	97	3.3	252	22	AAAB6150	Protein of the inv
7	94.5	3.3	471	21	AAV70274	Human purino recep
8	94.5	3.3	1714	21	AAAB18275	Plasmodium falcipa
9	94.5	3.2	379	21	AAAG12775	Arabidopsis thalia
10	94.5	3.2	383	21	AAAG12774	Arabidopsis thalia
11	94.5	3.2	395	22	AAAB5799	RpP5-like protein

12	93	3.2	404	21	AAV70275	Human purino recep
13	93	3.2	497	21	AAV70227	Human purino recep
14	91.5	3.1	592	11	AAAB05227	Sequence of rat C
15	90	3.1	253	21	AAAB43066	Human ORFX ORF2830
16	89	3.1	1792	21	AAAB46447	Mouse lamrin 8 po
17	89	3.1	1816	21	AAAB44446	Mouse lamrin 8 po
18	88.5	3.0	905	22	AAAB72198	Pepper B52 amino a
19	88.5	3.0	1004	21	AAAB18214	Plasmodium falcipa
20	88.5	3.0	2013	21	AAAB18265	Plasmodium falcipa
21	88.5	3.0	2954	20	AAV01632	Amino acid sequenc
22	88	3.0	839	21	AAAB56864	Human prostate can
23	87.5	3.0	739	15	AAAB49733	Sequence of the 1E
24	87	3.0	727	21	AAAG54101	Arabidopsis thalia
25	87	3.0	765	21	AAAG54100	Arabidopsis thalia
26	87	3.0	815	21	AAAG54099	Arabidopsis thalia
27	87	3.0	986	18	AAAB05930	Arabidopsis thalia
28	87	3.0	986	18	AAAB23552	Seq. ID 1 from W097
29	87	3.0	1164	18	AAAB23552	Cysteine loop prot
30	87	3.0	1164	21	AAAB01263	SCP449 peptidase (
31	86.5	3.0	739	11	AAAB08118	Vascular cell adhe
32	86.5	3.0	739	12	AAAB10316	Human VCAN-1. Hom
33	86.5	3.0	739	12	AAAB13906	Human VCAN-1. Hom
34	86.5	3.0	739	14	AAAB38548	Human VCAN-1. Hom
35	86.5	3.0	739	19	AAAB46734	Amino acid sequenc
36	86.5	3.0	744	21	AAV94408	Human VCAN-1/1651-
37	86.5	3.0	1073	18	AAAB32063	Human ST receptor
38	86.5	3.0	1073	19	AAAB37371	C. elegans alpha-1
39	86	3.0	355	21	AAV70408	Human purino recep
40	85.5	2.9	447	21	AAV70276	Human purino recep
41	85.5	2.9	739	13	AAAB21081	Sucrose synthase P
42	85.5	2.9	805	21	AAAB28141	Eucalyptus grandis
43	85.5	2.9	805	21	AAAB16282	Eucalyptus grandis
44	85.5	2.9	805	21	AAAB16336	Eucalyptus grandis
45	85.5	2.9	3079	15	AAAB39926	GAP protein Iira2.

ALIGNMENTS

RESULT 1

ID AAV6999 standard; Protein: 1237 AA.

AAV6999.

31-OCT-2000 (first entry)

S. cerevisiae essential gene YGR090W product.

YGR090W: essential gene, survival; growth; antifungal; inhibitor.

Saccharomyces cerevisiae.

MO200039342-A2.

06-JUL-2000.

29-DEC-1999; 99WO-0531042.

31-DEC-1998; 98US-0114734.

(MILL-) MILLENNIUM PHARM INC.

Bulawa C;

WP1: 2000-452421/39.

N-PSDB; AAAS1777.

Use of essential Saccharomyces cerevisiae genes and polypeptides for identifying antifungal compounds useful for treating fungal infections

Claim 36; Fig 3; 117pp; English.

DB 204 agsatswscspfkvcvciayfistdyrlv 233

RESULT 5

AAB67571
ID AAB67571 standard; Protein: 252 AA.

XX AAB67571;

XX 15-MAY-2001 (first entry)

XX Human PRO1327.

XX Human; PRO protein; mapping.

XX Homo sapiens.

XX WO200116318-A2.

XX 04-MAR-2001.

XX 24-AUG-2000; 2000WO-US23328.

XX 01-SEP-1999; 99WO-US20111.

XX 15-SEP-1999; 99WO-US21090.

XX 07-DEC-1999; 99US-0169495.

XX 04-DEC-1999; 99US-0170262.

XX 11-JAN-2000; 2000US-0175481.

XX 18-FEB-2000; 2000WO-US04342.

XX 22-FEB-2000; 2000WO-US04414.

XX 01-MAR-2000; 2000WO-US05601.

XX 03-MAR-2000; 2000US-0187202.

XX 25-APR-2000; 2000US-0199397.

XX 22-MAY-2000; 2000WO-US14042.

XX 05-JUN-2000; 2000US-0209832.

XX (GENTH) GENENTECH INC.

XX Eaton JL, Filvaroff E, Gerritsen ME, Goddard A, Godowski PJ;

XX Grimaldi CJ, Gurney AL, Watanabe CK, Wood WI;

XX WPI: 2001-183260/18.

XX N-PSDH, AAF92103.

XX Eighty four nucleic acids encoding PRO polypeptides, useful in

XX molecular biology, including use as hybridization probes, and in

XX chromosome and gene mapping.

XX Claim 12: Fig 92: 278pp; English.

XX The present sequence is a human PRO polypeptide (secreted and

XX transmembrane). The PRO protein, and PRO agonists, PRO antagonists or

XX anti-PRO antibodies are useful for preparation of a medicament useful in

XX the treatment of a condition which is responsive to the PRO protein,

XX agonists, antagonists or anti-PRO antibodies. The PRO protein may also be

XX employed as molecular weight markers for protein electrophoresis. The PRO

XX coding sequence has applications in molecular biology, including use as

XX hybridization probes, and in chromosome and gene mapping.

XX Sequence 252 AA:

Query Match 338; Score 97; DB 22; Length 252;

Best Local Similarity 23.38; Pred. No. 0.17; Mismatches 55; Indels 28; Gaps 6;

Matches 15; Conservative 32; Mismatches 55; Indels 28; Gaps 6;

DB 131 RKQYG-GDPLRRMSSPALTAGASGVMDFNNGTYLVSTFLFMEGQVSLILLIHPSEG-188

DB 97 KLIIGVQDLY-SNKTVALNLLVTKIVGHNGTFSVHFQHNATGQNLSTLSTLPPSKAV-155

DB 189 -----ASALWRARNGYDKIIFGKGFVNKTSNHFTECGTLTNSNAELCEYLD-235

DB 156 efhqeqqfleeakskifncrme-wek-vergr-----rslcthpakicardh 203

OY 236 DRDQEAIFYCMKPKQHMPCALTYMTTRNREV 265

DB 204 agsatswscspfkvcvciayfistdyrlv 233

RESULT 6

AAB6150
ID AAB6150 standard; protein: 252 AA.

XX AAB6150;

XX 02-APR-2001 (first entry)

XX Protein of the invention #62.

XX Secreted; transmembrane; gene therapy.

XX Unidentified.

XX WO200078961-A1.

XX 28-DEC-2000.

XX 18-FEB-2000; 2000WO-US04342.

XX 23-JUN-1999; 99US-0141037.

XX 20-JUL-1999; 99US-0144758.

XX 26-JUL-1999; 99US-0145698.

XX 01-SEP-1999; 99WO-US20111.

XX 29-OCT-1999; 99US-0162506.

XX 30-NOV-1999; 99WO-US28313.

XX 02-DEC-1999; 99WO-US28551.

XX 16-DEC-1999; 99WO-US30095.

XX 05-JAN-2000; 2000WO-US00219.

XX 06-JAN-2000; 2000WO-US00376.

XX (GENTH) GENENTECH INC.

XX Baker KP, Botstein D, Desnoyers L, Eaton DL, Ferrara N, Fong S;

XX Gao W, Goddard A, Godowski PJ, Grimaldi CJ, Gurney AL, Hillan KJ;

XX Pan J, Paoni NF, Roy MA, Smith V, Stewart TA, Tumas D;

XX Watanabe CK, Williams PM, Wood WI;

XX WPI: 2001-071395/08.

XX Secreted and transmembrane proteins and nucleic acids designated PRO,

XX useful as hybridization probes, in chromosome and gene mapping and gene

XX therapy.

XX Claim 1; Fig 124: 787pp; English.

XX The present invention relates to secreted and transmembrane proteins.

XX These proteins and the DNA encoding them may be used as hybridization

XX probes, in chromosome and gene mapping and in the generation of

XX anti-sense RNA and DNA. They may also be used used to generate either

XX transgenic animals or knockout animals which are in turn useful for

XX development and screening of therapeutically useful reagents.

XX The nucleic acids may also be used in gene therapy.

XX Sequence 252 AA:

Query Match 338; Score 97; DB 22; Length 252;

Best Local Similarity 23.38; Pred. No. 0.17; Mismatches 55; Indels 28; Gaps 6;

Matches 35; Conservative 32; Mismatches 55; Indels 28; Gaps 6;

DB 131 RKQYG-GDPLRRMSSPALTAGASGVMDFNNGTYLVSTFLFMEGQVSLILLIHPSEG-188

DB 97 KLIIGVQDLY-SNKTVALNLLVTKIVGHNGTFSVHFQHNATGQNLSTLSTLPPSKAV-155

DB 189 -----ASALWRARNGYDKIIFGKGFVNKTSNHFTECGTLTNSNAELCEYLD-235

```
Db      156 efiqsgqlfleakasklfcfrme~vek-verg-----tslctdpakiclrhn 205
QY      236 DRDFAFYCMKPMHPCCEALITYMTTRNEY 265
       : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      204 agqsatwscsqpfkvvcvylafstcdyrlv 233
```

RESULT 7
MAY70274

1D AAY70274 standard; protein; 471 AA.

AC AAY70274
xy

BT 20-JUN-2000 (first entry)
XX

Human purino receptor P2X₂a (hp2X₂aR) protein

KM Purino receptor; human; *hP2X₂*AR; ligand-gated ion channel; analgesic;
KM auditory; agonist; antagonist; modulator; pain; treatment; vestibular;
KM neuroendocrine disorder; gastrointestinal tract.
XX

OS Homo sapiens

PN WO2000011167-A2.
XY

PD 02-MAR-2000.

PF 20-AUG-1999; 99WO-US19042

PR 20-AUG-1998; 98US-0137458.
YY

PA (ABBO) ABBOTT LAB
XX

PL Lynch RJ, Burgard EC, Metzger RE, Niforatos W, Touma EB, XX

DR WPI; 2000-237649/20
DR N-PSDB: AA751321

DR N-PSDB; AA251321

PT Novel DNA encoding human P2X2 receptor, vectors and methods of
PT identifying modulators of the receptor, useful for relieving pain and
PT treating auditory, vestibular, neuroendocrine and gastrointestinal
PT tract disorders -

PS Claim 21; Flg 8A; 45pp; English

The present amino acid sequence is the human purino receptor P2X₂ (hP2X₂) protein, derived from human pituitary gland tissues. P2X₂ receptors are ligand-gated ion channels, activated by ATP and other natural or synthetic nucleotides. It has analgesic and auditory activity. P2X₂ sequences are useful for recombinant production and identification of therapeutic compounds, that modulate or interact with the P2X₂ receptor. P2X₂ agonists and antagonists are useful for relieving pain and treating neuroendocrine, auditory and vestibular, and gastrointestinal tract disorders.

sq Sequence 471 AA;

Query Match	3.38;	Score 96.5;	DB 21;	Length 471;
-------------	-------	-------------	--------	-------------

Matches 88; Conservative 46; Mismatches 120; Indels 157; Gaps 23;

```

OY      11 LLIILSSVTKMIFISFISNFKRLMSALNISIVHWNNKSASLEPKSII-PLAKLTET 63
       ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :
Db      50 lllllyf-vvvyyflavqsyge-----setgsesalltkvskltts 88
       ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :
OY      70 ELRKKEIIEKLDOQIPRP-----ETHVMTTSATSHT-----ATLMPRTYCR 11
       ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :
Db      89 ekkvwadveeyvk--ppeegavfsllrvteatbsgfgscpselrvmactlsdad--cv 14
       ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :
OY      115 GPDLIDILEVRHDLGQRKQYGDEFLRA-----RMSSPALTAGASC-----KWVDFFN 16
       ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :

```

```

Db 144 ageldml-----gnlrlrgreyyvyqpskvtvrlakcpodhasvst -1 144
OY 162 GYLVSLFLEEGVSLSLLIHPESASALMARNOGYDK-----LFF-GR 208
Db 189 gemaqfclilkn--slhyprfhfskgula--drtdykrctheasdyepclrlst 240
OY 209 FVNGTSVFTE-----GGLLSNMLEFVLDHDDGFAVY--KKVLMHME 250
Db 244 lvekgesftlakggvlgvlinwdcl-----dlpasecpkysfrldphvbas 290
OY 254 A-----LTYMTTN-----REVSYLTIKENS:FRKSVGY-E 294
Db 297 sgyntfakyyklgtctclikayrlgvlvghgaqstlplrltatalstspgqs 340
OY 285 MKKDKRHIDVTNCK--REKLEENQ-----VCMKPFVNG 317
Db 357 flcdalllttmkakyskklcdxvcpstpspswplarlavlgqapfepj 407

```

RESULT
AAB182

XX ID AAB182/3 Standard; Protein; 1/14 AA
XX

AC AAB18275;
XY

07-NOV-2000 (first entry)

DE Plasmodium falciparum chromosome 2 related protein SEQ ID NO:132.
XX

KM Plasmodium falciparum: chromosome 2; human malaria parasites; vaccine
KW antimalarial; malaria; protozoacides; infection; insecticide
XX

Plasmodium falciparum

PN W0200025728-A2

PD 11-MAY-2000
YX

PF 05-NOV-1999; 99WO-US26796.
VY

PR 05-NOV-1998: 98US-0107131.
YY

PA (HOFF/) HOFFMAN S.
PA (CARU/) CARUCCI D.
PA (GARD/) GARDNER M.
PA (VENT/) VENTER J C

PI Hoffman S, Carucci D, Gardner M, Venter J

DR WPI; 2000-365347/31.
yx

PT Proteins encoded by chromosome 2 of the human malarial parasite
PT Plasmodium falciparum, useful as antimalarial vaccines and in the
PT diagnosis of P.falciparum infection

PS Disclosure; Page 312-316; 577pp: English
 XX

CC The present invention describes proteins and their fragments (1) encoded
CC by chromosome 2 of the human malarial parasite, *Plasmodium falciparum*.
CC Also described are: (1) nucleotide sequences (II) encoding (I); and (2)
CC vaccines against *P. falciparum* infection comprising (I) or (II).
CC (I) and (II) are useful for the development of vaccines against
CC *P. falciparum* infection. (I) and polyclonal antisera or a monoclonal
CC antibody raised to immunogens comprising the sequences of (I), are
CC useful in the detection of infection with *P. falciparum*. Furthermore,
CC (I) (especially when they are rifins or secreted or membrane proteins)
CC can aid the identification of drugs to treat or prevent *P. falciparum*
CC infection, or they can be used to identify drug resistance in
CC *P. falciparum*. Sequencing of the plasmodium chromosome 2 and the
CC subsequent identification of proteins encoded by it will help to extend
CC our understanding of parasite biology, a process hampered by the
CC complexity of the parasite life cycle, and provide new targets for
CC vaccine and drug development. Parasite resistance to drugs and a slow

CC resistance to insecticides have led to a resurgence of malaria in many
CC parts of the world, and there is a pressing need for vaccines and new
CC drugs. AAA70078 to AAA70287 and AAB18144 to AAB18352 represent nucleotide
CC and protein sequences given in the present invention, but which are not
CC specifically mentioned within the specification.

XX Sequence 1714 AA:

Query Match 33%; Score 95; DB 21; Length 1714;
Best Local Similarity 19.8%; Pred. No. 6.6; Mismatches 82; Gaps 15;
Matches 61; Conservative 53;

OY 225 NSMAELCEYLDDFDGFAFYC-----MKPQHPCEALTYWTRNREVS-----L 268
DB 1107 nmndlnhdnndnngfnsceclnlgpveqkcegeqektkkkylynnuykkgknei 1166
OY 269 TDKENSILFHSKSYGVEM--MKDRKHIDVTCNKRKEIEETQVC--MKPPYCGTTLQCK 324
DB 1167 hddynlksbstlnynlnlnkdkh-----nmndgkscelkcsipyvkekynlen- 1219
OY 345 WITFCNVVLDTKINCLCKLIVLDGDTLRQWYFFKVVKTLEF-----DLHET 379
DB 1220 -----pyletfg-----llygdksgyykciomnkkyamkvylkecheitvd 1263
OY 480 GIFKKHLLDAERHTQIQMKKHSYFVTQYLSLI--DHDYIPREIDRLSGDKNTAVITF 438
DB 1464 nllkkyllfknpkhnll-----lslydlfcmnyiclimdyceg-----stllidy 1308
OY 439 GQHFREPTDIFIRRAICVOKAIEFLRSP---ATKYI---IKTENI-----REWHI 485
DB 1309 fmslvpsidvy-----elkkimknllaldfhsmnlhndklenimfkknkkkkrf 1363
OY 446 ETEFGDP 493
DB 1364 nyekygst 1371

RESULT 9
AAG12775
ID AAG12775 standard: protein; 379 AA.

AC AAG12775;

DT 17-OCT-2000 (first entry)

tr Arabidopsis thaliana protein fragment seq ID NO: 12018.

XX Protein identification: signal transduction pathway; metabolic pathway;
KM hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.

XX Arabidopsis thaliana.

PN EPI013405-A2.

PO 06-SEP-2000.

FF 25-FEB-2000; 2000FP-0301439.

XX 25-FEB-1999; 9905-0121825.
PR 05-MAR-1999; 9905-0123180.
PP 09-MAR-1999; 9905-0123548.
PR 21-MAR-1999; 9905-0125788.
PR 25-MAR-1999; 9905-0126264.
PR 29-MAR-1999; 9905-0126785.
PR 01-APR-1999; 9905-0127462.
PR 06-APR-1999; 9905-0128234.
PR 08-APR-1999; 9905-0128714.
PR 14-APR-1999; 9905-0129845.
PR 19-APR-1999; 9905-0130077.
PR 21-APR-1999; 9905-0130449.
PR 23-APR-1999; 9905-0130510.

PR 23-APR-1999; 9905-0130891.
PR 28-APR-1999; 9905-0131449.
PR 30-APR-1999; 9905-0132048.
PR 04-APR-1999; 9905-0132407.
PR 10-MAY-1999; 9905-0132484.
PR 05-MAY-1999; 9905-0132485.
PR 06-MAY-1999; 9905-0132486.
PR 06-MAY-1999; 9905-0132487.
PR 07-MAY-1999; 9905-0132488.
PR 11-MAY-1999; 9905-0134256.
PR 14-MAY-1999; 9905-0134218.
PR 14-MAY-1999; 9905-0134219.
PR 14-MAY-1999; 9905-0134221.
PR 18-MAY-1999; 9905-0134370.
PR 19-MAY-1999; 9905-0134768.
PR 20-MAY-1999; 9905-0134941.
PR 21-MAY-1999; 9905-0135124.
PR 24-MAY-1999; 9905-0135353.
PR 25-MAY-1999; 9905-0135629.
PR 27-MAY-1999; 9905-0136021.
PR 28-MAY-1999; 9905-0136392.
PR 01-JUN-1999; 9905-0137222.
PR 03-JUN-1999; 9905-0137528.
PR 04-JUN-1999; 9905-0137502.
PR 07-JUN-1999; 9905-0137724.
PR 08-JUN-1999; 9905-0138094.
PR 10-JUN-1999; 9905-0138540.
PR 14-JUN-1999; 9905-0138847.
PR 16-JUN-1999; 9905-0139119.
PR 16-JUN-1999; 9905-0139412.
PR 17-JUN-1999; 9905-0139453.
PR 18-JUN-1999; 9905-0139492.
PR 18-JUN-1999; 9905-0139454.
PR 18-JUN-1999; 9905-0139455.
PR 18-JUN-1999; 9905-0139456.
PR 18-JUN-1999; 9905-0139457.
PR 18-JUN-1999; 9905-0139458.
PR 18-JUN-1999; 9905-0139459.
PR 18-JUN-1999; 9905-0139460.
PR 18-JUN-1999; 9905-0139461.
PR 18-JUN-1999; 9905-0139462.
PR 18-JUN-1999; 9905-0139463.
PR 18-JUN-1999; 9905-0139750.
PR 18-JUN-1999; 9905-0139763.
PR 21-JUN-1999; 9905-0139817.
PR 22-JUN-1999; 9905-0139859.
PR 23-JUN-1999; 9905-0140353.
PR 23-JUN-1999; 9905-0140354.
PR 24-JUN-1999; 9905-0140695.
PR 28-JUN-1999; 9905-0140823.
PR 29-JUN-1999; 9905-0140991.
PR 30-JUN-1999; 9905-0141287.
PR 01-JUL-1999; 9905-0141842.
PR 01-JUL-1999; 9905-0142154.
PR 02-JUL-1999; 9905-0142055.
PR 06-JUL-1999; 9905-0142390.
PR 08-JUL-1999; 9905-0142803.
PR 09-JUL-1999; 9905-0142920.
PR 12-JUL-1999; 9905-0142977.
PR 13-JUL-1999; 9905-0143542.
PR 14-JUL-1999; 9905-0143624.
PR 15-JUL-1999; 9905-0144005.
PR 16-JUL-1999; 9905-0144005.
PR 16-JUL-1999; 9905-0144086.
PR 19-JUL-1999; 9905-0144325.
PR 19-JUL-1999; 9905-0144331.
PR 19-JUL-1999; 9905-0144332.
PR 19-JUL-1999; 9905-0144333.
PR 19-JUL-1999; 9905-0144334.
PR 19-JUL-1999; 9905-0144335.
PR 20-JUL-1999; 9905-0144352.
PR 20-JUL-1999; 9905-0144632.

PR	14-OCT-1999	9905-0155638.
PR	18-OCT-1999	9905-0159584.
PR	21-OCT-1999	9905-0160741.
PR	21-OCT-1999	9905-0160767.
PR	21-OCT-1999	9905-0160768.
PR	21-OCT-1999	9905-0160770.
PR	21-OCT-1999	9905-0160814.
PR	21-OCT-1999	9905-0160815.
PR	22-OCT-1999	9905-0160980.
PR	22-OCT-1999	9905-0160981.
PR	22-OCT-1999	9905-0160989.
PR	23-OCT-1999	9905-0161404.
PR	25-OCT-1999	9905-0161405.
PR	25-OCT-1999	9905-0161406.
PR	26-OCT-1999	9905-0161359.
PR	26-OCT-1999	9905-0161360.
PR	28-OCT-1999	9905-0161920.
PR	28-OCT-1999	9905-0161992.
PR	28-OCT-1999	9905-0161993.
PR	29-OCT-1999	9905-0162142.

Best Local Similarity	20.5%	Pred No	0.61	Length	379
Very match	3.28	Score	94.5	DB	21

Matches	53;	Conservative	31;	Mismatches	72;	Indels	103;	Gaps	10
---------	-----	--------------	-----	------------	-----	--------	------	------	----

-----MKPPVGGYTLGCKWITFCNQ 332

QY 333 VQD1KINGCLKGLIYIIGDSTI.B-(W-----

103 pftdalkflgkmrgkrimlvgsdmnrnqwsllvclvqsvlptlrrkkltyngp---clmsl 158

Db 159 hsldefetsiefcwaplivelkrgvdrkvlhldslednarywrvqvdvlfdsahwvlhsa 218

[illegible]

QY 464 LEIRSPATKVITKTENIRF A83

100

AA612774

standard; protein; 383 AA.

AC
XX

XX Arabidopsis thaliana

Protein 100kDa: 11
XX XX

Fragment size in NO₂-12017

termination sequence.

05 Arabidopsis thaliana.

PN EP1033405-A2.
XX

XX	25-FEB-2000:
PF	

25-FEB-1999; 99US-0121825

09-MAR-1999; 99US-0123548

20 - MAR - 1999; 9905-0125788.

PR	25-MAR-1999;	99US-0126264.	PR	16-JUL-1999;	99US-0144086.
PR	24-MAR-1999;	99US-0126785.	PR	19-JUL-1999;	99US-0144325.
PR	01-APR-1999;	99US-0127462.	PR	19-JUL-1999;	99US-0144331.
PR	06-APR-1999;	99US-0128234.	PR	19-JUL-1999;	99US-0144337.
PR	08-APR-1999;	99US-0128714.	PR	19-JUL-1999;	99US-0144333.
PR	16-APR-1999;	99US-0129845.	PR	19-JUL-1999;	99US-0144334.
PR	19-APR-1999;	99US-0130077.	PR	19-JUL-1999;	99US-0144335.
PR	21-APR-1999;	99US-0130449.	PR	20-JUL-1999;	99US-0144352.
PR	24-APR-1999;	99US-0130510.	PR	20-JUL-1999;	99US-0144632.
PR	23-APR-1999;	99US-0130891.	PR	20-JUL-1999;	99US-0144684.
PR	28-APR-1999;	99US-0131449.	PR	21-JUL-1999;	99US-0144814.
PR	30-APR-1999;	99US-0132048.	PR	21-JUL-1999;	99US-0145086.
PR	30-APR-1999;	99US-0132407.	PR	21-JUL-1999;	99US-0145088.
PR	04-MAY-1999;	99US-0132484.	PR	22-JUL-1999;	99US-0145085.
PR	05-MAY-1999;	99US-0132485.	PR	22-JUL-1999;	99US-0145087.
PR	06-MAY-1999;	99US-0132486.	PR	22-JUL-1999;	99US-0145089.
PR	06-MAY-1999;	99US-0132487.	PR	22-JUL-1999;	99US-0145192.
PR	07-MAY-1999;	99US-0132853.	PR	23-JUL-1999;	99US-0145145.
PR	11-MAY-1999;	99US-0134256.	PR	23-JUL-1999;	99US-0145218.
PR	14-MAY-1999;	99US-0134219.	PR	23-JUL-1999;	99US-0145224.
PR	14-MAY-1999;	99US-0134218.	PR	26-JUL-1999;	99US-0145276.
PR	14-MAY-1999;	99US-0134221.	PR	27-JUL-1999;	99US-0145913.
PR	14-MAY-1999;	99US-0134370.	PR	27-JUL-1999;	99US-0145918.
PR	14-MAY-1999;	99US-0134768.	PR	27-JUL-1999;	99US-0145919.
PR	19-MAY-1999;	99US-0134941.	PR	28-JUL-1999;	99US-0145951.
PR	20-MAY-1999;	99US-0135124.	PR	02-AUG-1999;	99US-0146386.
PR	21-MAY-1999;	99US-0135353.	PR	02-AUG-1999;	99US-0146388.
PR	24-MAY-1999;	99US-0136021.	PR	03-AUG-1999;	99US-0146389.
PR	25-MAY-1999;	99US-0136392.	PR	03-AUG-1999;	99US-0147038.
PR	27-MAY-1999;	99US-0136782.	PR	04-AUG-1999;	99US-0147204.
PR	28-MAY-1999;	99US-0137222.	PR	04-AUG-1999;	99US-0147302.
PR	01-JUN-1999;	99US-0137528.	PR	05-AUG-1999;	99US-0147192.
PR	03-JUN-1999;	99US-0137502.	PR	05-AUG-1999;	99US-0147260.
PR	04-JUN-1999;	99US-0137724.	PR	06-AUG-1999;	99US-0147303.
PR	07-JUN-1999;	99US-0138094.	PR	06-AUG-1999;	99US-0147403.
PR	08-JUN-1999;	99US-0138540.	PR	09-AUG-1999;	99US-0147935.
PR	10-JUN-1999;	99US-0139119.	PR	10-AUG-1999;	99US-0148171.
PR	10-JUN-1999;	99US-0139457.	PR	11-AUG-1999;	99US-0148319.
PR	14-JUN-1999;	99US-0139453.	PR	12-AUG-1999;	99US-0148341.
PR	16-JUN-1999;	99US-0139456.	PR	13-AUG-1999;	99US-0148565.
PR	17-JUN-1999;	99US-0139454.	PR	13-AUG-1999;	99US-0148684.
PR	18-JUN-1999;	99US-0139455.	PR	16-AUG-1999;	99US-0149368.
PR	18-JUN-1999;	99US-0139457.	PR	17-AUG-1999;	99US-0149175.
PR	18-JUN-1999;	99US-0139458.	PR	18-AUG-1999;	99US-0149426.
PR	18-JUN-1999;	99US-0139459.	PR	20-AUG-1999;	99US-0149723.
PR	18-JUN-1999;	99US-0139460.	PR	20-AUG-1999;	99US-0149727.
PR	18-JUN-1999;	99US-0139461.	PR	23-AUG-1999;	99US-0149929.
PR	18-JUN-1999;	99US-0139462.	PR	23-AUG-1999;	99US-0149902.
PR	18-JUN-1999;	99US-0139463.	PR	25-AUG-1999;	99US-0149930.
PR	18-JUN-1999;	99US-0139750.	PR	25-AUG-1999;	99US-0150566.
PR	18-JUN-1999;	99US-0139763.	PR	26-AUG-1999;	99US-0150684.
PR	21-JUN-1999;	99US-0139817.	PR	27-AUG-1999;	99US-0151065.
PR	22-JUN-1999;	99US-0139859.	PR	27-AUG-1999;	99US-0151066.
PR	23-JUN-1999;	99US-0140353.	PR	30-AUG-1999;	99US-0151303.
PR	24-JUN-1999;	99US-0140695.	PR	31-AUG-1999;	99US-0151438.
PR	28-JUN-1999;	99US-0140823.	PR	01-SEP-1999;	99US-0151930.
PR	29-JUN-1999;	99US-0140991.	PR	07-SEP-1999;	99US-0152363.
PR	30-JUN-1999;	99US-0141287.	PR	10-SEP-1999;	99US-0153070.
PR	01-JUL-1999;	99US-0141642.	PR	13-SEP-1999;	99US-0153758.
PR	01-JUL-1999;	99US-0142154.	PR	15-SEP-1999;	99US-0154018.
PR	02-JUL-1999;	99US-0142055.	PR	16-SEP-1999;	99US-0154779.
PR	04-JUL-1999;	99US-0142390.	PR	20-SEP-1999;	99US-0154779.
PR	04-JUL-1999;	99US-0142803.	PR	22-SEP-1999;	99US-0155139.
PR	09-JUL-1999;	99US-0142920.	PR	24-SEP-1999;	99US-0155486.
PR	12-JUL-1999;	99US-0143287.	PR	24-SEP-1999;	99US-0155659.
PR	13-JUL-1999;	99US-0143542.	PR	28-SEP-1999;	99US-0156458.
PR	14-JUL-1999;	99US-0143624.	PR	29-SEP-1999;	99US-0156596.
PR	15-JUL-1999;	99US-0144005.	PR	04-OCT-1999;	99US-0157717.
PR	16-JUL-1999;	99US-0144085.	PR	05-OCT-1999;	99US-0157753.
			PR	06-OCT-1999;	99US-0157865.
			PR	07-OCT-1999;	99US-0158029.


```

XX      W0200011167-A2.
XX      02-MAR-2000.
XX      20-AUG-1999; 99MO-US19042.
XX      20-AUG-1998; 98US-0137458.
XX      (ABBO ) ABBOTT LAB.
XX      Lynch KJ, Burgard EC, Metzger RE, Niforatos W, Touma EB;
XX      WPI: 2000-237649/20.
XX      N-PSDB: AA251322.
XX
XX      Novel DNA encoding human P2X2 receptor, vectors and methods of
XX      identifying modulators of the receptor, useful for relieving pain and
XX      treating auditory, vestibular, neuroendocrine and gastrointestinal
XX      tract disorders.
XX      Claim 21; Fig 8B; 45pp; English.
XX
XX      The present amino acid sequence is the human purino receptor P2X2b
XX      (hP2X2b) protein, derived from human pituitary gland tissues. P2X2
XX      receptors are ligand-gated ion channels, activated by ATP and other
XX      natural or synthetic nucleotides. It has analgesic and auditory activity.
XX      P2X2b sequences are useful for recombinant production and identification
XX      of therapeutic compounds, that modulate or interact with the P2X2
XX      receptor. P2X2b agonists and antagonists are useful for relieving pain
XX      and treating neuroendocrine, auditory and vestibular, and
XX      gastrointestinal tract disorders.
XX
XX      Sequence 404 AA:
XX
XX      Query Match 3.2%; Score 93; DB 21; Length 404;
XX      Best Local Similarity 22.9%; Pred. No. 0.98;
XX      Matches 69; Conservative 36; Mismatches 86; Indels 110; Gaps 18;
XX
XX      11 LLLISFSVVTMMIFISQNTKLSALNLSISVHYMNSAKSLFPTSLI-PLKPILET 69
XX      ||||| : : : : : : : : : : : : : : : : : : : : : : : : :
XX      50 LLLIlyl---vwyvllvgksyqe-----setgpesslltkvvgilts 88
XX
XX      70 ELRIKEIIEKLDQIIPRP-----FTHVNTTTSATHST-----ATILNPDTRYCR 114
XX      : : : : : : : : : : : : : : : : : : : : : : : : :
XX      89 ehkxwvdeeyvk---ppegsgvslsltrveathtsgtgcpceslrvnmatclsdad--cv 143
XX
XX      115 GDOLDILLEVRHLOGRKQYGGDFLRA-----RMSSPALTLTGASG-----KYMDNN 161
XX      : : : : : : : : : : : : : : : : : : : : : : : : :
XX      144 ageidml-----gnglirgrcvpyyqpskcevfqwcpevdgasvsqf-1 188
XX
XX      162 GTLYVSFTLFWEGVSLSLILHPSEGASALMRANQYDK-----IIFK-GK 208
XX      || : : : : : : : : : : : : : : : : : : : : : : : : :
XX      189 gtmaphntlllkn--slhyprfkfhsqnia--dtdgylkrctfheasdlvcplfikg 243
XX
XX      209 FVNGTSHVFTE-----CGTLNSNAELCEYIDRDDEAFYC-----MKPOHMCE 253
XX      || : : : : : : : : : : : : : : : : : : : : : : : : :
XX      244 lvekgesftelahnkgvlyvlinwdcl-----dlpasecmypsftrldphvpas 296
XX
XX      297 s 297
XX
XX      RESULT 13
XX      AAAY70277
XX      10 AAAY70277 standard; Protein: 497 AA.
XX
XX      AAAY70277:
XX
XX      20-JUN-2000 (first entry)
XX

```

```

DE      Human purino receptor P2X2d (hP2X2dR) protein.
XX
XX      Purino receptor; human; hP2X2dR; ligand-gated ion channel; analgesic;
XX      auditory; agonist; antagonist; modulator; pain; treatment; vestibular;
XX      neuroendocrine disorder; gastrointestinal tract.
XX
XX      Homo sapiens.
XX      W0200011167-A2.
XX      02-MAR-2000.
XX      20-AUG-1999; 99MO-US19042.
XX      20-AUG-1998; 98US-0137458.
XX      (ABBO ) ABBOTT LAB.
XX      Lynch KJ, Burgard EC, Metzger RE, Niforatos W, Touma EB;
XX      WPI: 2000-237649/20.
XX      N-PSDB: AA251322.
XX
XX      Novel DNA encoding human P2X2 receptor, vectors and methods of
XX      identifying modulators of the receptor, useful for relieving pain and
XX      treating auditory, vestibular, neuroendocrine and gastrointestinal
XX      tract disorders.
XX      Claim 21; Fig 8D; 45pp; English.
XX
XX      The present amino acid sequence is the human purino receptor P2X2d
XX      (hP2X2dR) protein, derived from human pituitary gland tissues. P2X2
XX      receptors are ligand-gated ion channels, activated by ATP and other
XX      natural or synthetic nucleotides. It has analgesic and auditory activity.
XX      P2X2d sequences are useful for recombinant production and identification
XX      of therapeutic compounds, that modulate or interact with the P2X2
XX      receptor. P2X2d agonists and antagonists are useful for relieving pain
XX      and treating neuroendocrine, auditory and vestibular, and
XX      gastrointestinal tract disorders.
XX
XX      Sequence 497 AA:
XX
XX      Query Match 3.2%; Score 93; DB 21; Length 497;
XX      Best Local Similarity 22.9%; Pred. No. 1.4;
XX      Matches 69; Conservative 36; Mismatches 86; Indels 110; Gaps 18;
XX
XX      11 LLLISFSVVTMMIFISQNTKLSALNLSISVHYMNSAKSLFPTSLI-PLKPILET 69
XX      ||||| : : : : : : : : : : : : : : : : : : : : : : : : :
XX      50 LLLIlyl---vwyvllvgksyqe-----setgpesslltkvvgilts 88
XX
XX      70 ELRIKEIIEKLDQIIPRP-----FTHVNTTTSATHST-----ATILNPDTRYCR 114
XX      : : : : : : : : : : : : : : : : : : : : : : : : :
XX      89 ehkxwvdeeyvk---ppegsgvslsltrveathtsgtgcpceslrvnmatclsdad--cv 143
XX
XX      115 GDOLDILLEVRHLOGRKQYGGDFLRA-----RMSSPALTLTGASG-----KYMDNN 161
XX      : : : : : : : : : : : : : : : : : : : : : : : : :
XX      144 ageidml-----gnglirgrcvpyyqpskcevfqwcpevdgasvsqf-1 188
XX
XX      162 GTLYVSFTLFWEGVSLSLILHPSEGASALMRANQYDK-----IIFK-GK 208
XX      || : : : : : : : : : : : : : : : : : : : : : : : : :
XX      189 gtmaphntlllkn--slhyprfkfhsqnia--dtdgylkrctfheasdlvcplfikg 243
XX
XX      209 FVNGTSHVFTE-----CGTLNSNAELCEYIDRDDEAFYC-----MKPOHMCE 253
XX      || : : : : : : : : : : : : : : : : : : : : : : : : :
XX      244 lvekgesftelahnkgvlyvlinwdcl-----dlpasecmypsftrldphvpas 296
XX
XX      297 s 297
XX
XX      RESULT 14
XX

```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 12, 2001, 16:03:51 : Search time 20.85 seconds

(without alignments)
540.186 Million cell updates/sec

Title: US-09-729-454-2
Perfect score: 2908
Sequence: 1 MSSNMLQKTLILISFSVY.....IHPDHYGNQIMFLNTIC 547

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 197339 seqs, 20590346 residues

Total number of hits satisfying chosen parameters: 197339

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: Issued Patents AA: *
2: /cgn2_6/prodata/2/1aa/5A.COMB.pep: *
3: /cgn2_6/prodata/2/1aa/6A.COMB.pep: *
4: /cgn2_6/prodata/2/1aa/6B.COMB.pep: *
5: /cgn2_6/prodata/2/1aa/PCUS.COMB.pep: *
6: /cgn2_6/prodata/2/1aa/backfillseq1.pep: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	98	3.4	698	1 US-08-376-362A-20	Sequence 20, Appl
2	96.5	3.3	471	4 US-09-191-608-17	Sequence 17, Appl
3	94	3.2	404	4 US-09-191-608-18	Sequence 18, Appl
4	93	3.2	497	4 US-09-191-608-20	Sequence 20, Appl
5	90.5	3.1	582	4 US-08-194-560-2	Sequence 2, Appl
6	87	3.0	1164	2 US-08-589-756-1	Sequence 9, Appl
7	84.5	2.9	739	5 PCT-US93-00031-9	Sequence 1, Appl
8	86	3.0	1494	3 US-08-589-587-186	Sequence 186, Appl
9	85.5	2.9	447	4 US-09-191-608-19	Sequence 19, Appl
10	85.5	2.9	884	6 5208144-8	Sequence 4, Appl
11	85.5	2.9	3079	5 PCT-US94-00198-4	Sequence 2, Appl
12	85.5	2.9	3135	1 US-08-323-1708-2	Sequence 6, Appl
13	85	2.9	450	2 US-08-818-514-6	Sequence 8, Appl
14	85	2.9	450	4 US-09-115-934A-6	Sequence 8, Appl
15	85	2.9	766	2 US-08-553-436A-8	Sequence 2, Appl
16	85	2.9	2516	3 US-08-374-077C-2	Sequence 2, Appl
17	85	2.9	2516	4 US-08-895-590-2	Sequence 3, Appl
18	84.5	2.9	1150	2 US-08-589-756-3	Sequence 2, Appl
19	84.5	2.9	3418	2 US-08-639-501-2	Sequence 2, Appl
20	84.5	2.9	3418	3 US-09-044-946-2	Sequence 2, Appl
21	84.5	2.9	3418	3 US-09-044-908-2	Sequence 2, Appl
22	83.5	2.9	318	3 US-09-188-579-81	Sequence 81, Appl
23	83.5	2.9	318	4 US-09-315-444-81	Sequence 46, Appl
24	83.5	2.9	324	2 US-08-528-057-46	Sequence 43, Appl
25	83.5	2.9	1152	5 PCT-US96-01314-43	Sequence 43, Appl
26	83.5	2.9	1152	6 5424399-2	Sequence 43, Appl
27	83.5	2.9	1152	6 5424399-2	Sequence 43, Appl

ALIGNMENTS

28	83.5	2.9	1153	1 US-08-173-497-3	Sequence 3, Appl
29	83.5	2.9	1153	1 US-08-286-889-3	Sequence 3, Appl
30	83.5	2.9	1153	1 US-08-485-618-3	Sequence 3, Appl
31	83.5	2.9	1153	1 US-08-362-652-3	Sequence 3, Appl
32	83.5	2.9	1153	2 US-08-605-672-3	Sequence 3, Appl
33	83.5	2.9	1153	3 US-08-482-293A-3	Sequence 3, Appl
34	83.5	2.9	1153	2 US-08-943-363-3	Sequence 3, Appl
35	83.5	2.9	1153	4 US-09-193-043-3	Sequence 49, Appl
36	82.5	2.8	1050	3 US-09-045-632-49	Sequence 50, Appl
37	82.5	2.8	1050	3 US-09-045-632-50	Sequence 50, Appl
38	82.5	2.8	1167	2 US-08-589-756-2	Sequence 9, Appl
39	82	2.8	834	2 US-08-677-734A-9	Sequence 10, Appl
40	82	2.8	2329	3 US-08-755-587-16	Sequence 16, Appl
41	82	2.8	2329	1 US-07-623-033-2	Sequence 2, Appl
42	79.5	2.7	1075	1 US-08-603-753D-4	Sequence 4, Appl
43	79.5	2.7	3418	3 US-08-755-587-44	Sequence 4, Appl
44	79.5	2.7	3418	4 US-09-099-753-4	Sequence 4, Appl
45	79.5	2.7	3418	4 US-09-099-753-4	Sequence 4, Appl

RESULT 1
US-08-376-362A-20
Sequence 20, Application US/08376362A
Patent No. 5693756
GENERAL INFORMATION:
APPLICANT: Li, Xiao-Jiang
APPLICANT: Snyder, Solomon H.
TITLE OF INVENTION: AMILORIDE-SENSITIVE SODIUM CHANNEL AND
TITLE OF INVENTION: METHOD OF IDENTIFYING SUBSTANCES WHICH STIMULATE OR BLOCK
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: Banner & Allegretti, LTD
STREET: 1001 G Street, N.W., Eleventh floor
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20001-4597
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/376,362A
FILING DATE: 23-JAN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Kagan A., Sarah
REGISTRATION NUMBER: 32,141
REFERENCE/DOCKET NUMBER: 01107,48125
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-508-9100
TELEFAX: 202-508-9299
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 698 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHEICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Rattus rattus
US-08-376-362A-20
Query Match 3.4%; Score 98; DB 1; Length 698;
Best Local Similarity 19.2%; Pred. No. 0.28;

Matches 89: Conservative 53; Mismatches 157; Indels 164; Gaps 18:

```

QY 11 LLLISFSVVTWMIFFISQNTFKLMSALNLSISVHYNNNSAKSLFPKSLPLKPLTE 70
DB 114 VLMLCTFGMYWQFALLFEELSYSPVSLNLT-----NSDKLVFPAVTCILNPRTE 167
QY 71 LRIKEIEIKLDO-----QIPPRPF 89
DB 168 ---IKEELEEDRLTEQTLFDLKYKNSSTYQAGARRSSFDLLCAPRPLQRLKTPPPY 225
QY 90 THVNTTTSATHTATLNR-----DTYCGDQDLLEVRDH 127
DB 226 S---GRIARSSSSVSDNNPFQVDRKDMKIGFQLCNKKSDCFYQYSSG---YDAVRE---- 277
QY 128 LGQRKQYGDPLFARMSPALTAGASGVNDPNNGYLVSFTLWEGQVSLSLIHSE 187
DB 278 -WYRFHYINLSLSDTSPALEEALG-----NFIICFRNAPCAQAYTS-KFHHNMY 329
QY 188 G-----ASALMPARNQGYDKIIFKGFVNGTSHVTECGTLTNSNAELCEYLDDRD 238
DB 330 GNCYTFPDKNNNSMLWMSMPG-----VNN-----GLSLTLRTQNDPI---- 367
QY 239 QEAFCMKPQHMPCALTYMTTRNREVSYLDEKNSLFHRSKVGVEMMDRKHIDVTNCN 298
DB 368 -----PLSLTVGARVWVGQDEPAFMDQGNLAPVETSISKREA--LDSLGN 417
QY 299 KREKIEETCOYKMKPRVPGGTTLOGKWTTFPCNOVDLTINIGCLKGLIYLLDSTLR 358
DB 418 YGDCYENGSDVPYKMLYPSKYTQO-----VCIHSCQEMNKKCGCA--- 459
QY 359 QMIYY-FPKVVKTLKF-----FDLHETGIRPK 384
DB 460 -YIFYPKKGVCFCDYRKQSMGYCYKLGACAFSLDSLCCPSK 501

RESULT 2
US-09-191-608-17
Sequence 17, Application US/09191608
Patent No. 6242216
GENERAL INFORMATION:
APPLICANT: Lynch, Kevin J.
APPLICANT: Burgard, Edward C.
APPLICANT: Metzger, Randy E.
APPLICANT: Nitrosatos, Wende
APPLICANT: Touma, Edward B.
TITLE OF INVENTION: Nucleic Acids Encoding a Functional
TITLE OF INVENTION: Human Purinoreceptor P2X2 and P2X4 And Methods of Production
FILE REFERENCE: 6394 US.P1
CURRENT APPLICATION NUMBER: US/09/191.608
CURRENT FILING DATE: 1998-11-13
NUMBER OF SEQ ID NOS: 26
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 17
LENGTH: 471
TYPE: PRT
ORGANISM: Homo sapiens
US-09-191-608-17

Query Match 3.38; Score 96.5; DB 4; Length 471:
Best Local Similarity 21.4%; Pred. No. 0.2;
Matches 88: Conservative 46; Mismatches 120; Indels 157; Gaps 23:
QY 11 LLLISFSVVTWMIFFISQNTFKLMSALNLSISVHYNNNSAKSLFPKSLPLKPLTE 69
DB 50 LLLILYF---VWYFIVOKSYOE-----STGPRESIITKKKGLTIS 88
QY 70 ELRIKEIEIKLDOQIPPRP-----FTHVNTTTSATHT-----ATLNRDPTGR 114
DB 89 EKKVMDVEEYK---PREGGSVFSIITRYEATHSQTOGTCPESIRVHNATCLSDAD--CV 143

```

```

QY 115 GEQDLLEVRDHIAQRKQYGGDFLRA-----KMSSPALFANAG-----KWMDFNR 17
DB 144 AGEIDLML-----GNGLRIGRCVPYQGPSTCTEFGWCTWEGASVSLFL 186
QY 162 GTYLVSTFLWEGQVSLSLIHSEGASALMARNOGYK-----IFK-GK 208
DB 189 GTMAPNFTILIKN--SIHYPKFHSKGNIA-----DRIDQYLKCTFHASGLVPIPIKLCF 244
QY 209 FVNGTSHVTE-----GGLTNSNAELCEYLDRODEARYC-----KNIYHMCCE 275
DB 244 IVEKAGESTELAHKGVIGVITIMDDCL-----DLPASECNKYSFRLPLPVPVPS 296
QY 254 A-----LTWETTRN-----REVSYLDEKNSLFHRSKVNYF 474
DB 297 SGVYFRKAKYKINGTTTTLIKAYGIRIDVIHGOAGKFSLLPTIINLATALISVGVGS 456
QY 285 MAKDRKHIDVTNCNK---REKIEETCO-----YGMKPRVPG 517
DB 357 FLCDWILLTFMKNKNVYSHKKKFDKVCIPSPHSGSNPVTLANVLSLAPPRG 407

RESULT 3
US-09-191-608-18
Sequence 18, Application US/09191608
Patent No. 6242216
GENERAL INFORMATION:
APPLICANT: Lynch, Kevin J.
APPLICANT: Burgard, Edward C.
APPLICANT: Metzger, Randy E.
APPLICANT: Nitrosatos, Wende
APPLICANT: Touma, Edward B.
TITLE OF INVENTION: Nucleic Acids Encoding a Functional
TITLE OF INVENTION: Human Purinoreceptor P2X2 and P2X4 And Methods of Production
FILE REFERENCE: 6394 US.P1
CURRENT APPLICATION NUMBER: US/09/191.608
CURRENT FILING DATE: 1998-11-13
NUMBER OF SEQ ID NOS: 26
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 18
LENGTH: 404
TYPE: PRT
ORGANISM: Homo sapiens
US-09-191-608-18

Query Match 3.28; Score 94; DB 4; Length 404:
Best Local Similarity 22.9%; Pred. No. 0.37;
Matches 69: Conservative 36; Mismatches 86; Indels 15; Gaps 17
QY 11 LLLISFSVVTWMIFFISQNTFKLMSALNLSISVHYNNNSAKSLFPKSLPLKPLTE 69
DB 50 LLLILYF---VWYFIVOKSYOE-----STGPRESIITKKKGLTIS 88
QY 70 ELRIKEIEIKLDOQIPPRP-----FTHVNTTTSATHT-----ATLNRDPTGR 114
DB 89 EKKVMDVEEYK---PREGGSVFSIITRYEATHSQTOGTCPESIRVHNATCLSDAD--CV 143
QY 115 GQDLDLLEVRDHIAQRKQYGGDFLRA-----KMSSPALFANAG-----KNIYHMCCE 191
DB 144 AGEIDLML-----GNGLRIGRCVPYQGPSTCTEFGWCTWEGASVSLFL 186
QY 162 GTYLVSTFLWEGQVSLSLIHSEGASALMARNOGYK-----IFK-GK 208
DB 189 GTMAPNFTILIKN--SIHYPKFHSKGNIA-----DRIDQYLKCTFHASGLVPIPIKLCF 244
QY 209 FVNGTSHVTE-----GGLTNSNAELCEYLDRODEARYC-----KNIYHMCCE 275
DB 244 IVEKAGESTELAHKGVIGVITIMDDCL-----DLPASECNKYSFRLPLPVPVPS 296
QY 254 A 254

```

Db 297 S 297

```

RESULT 4
US-09-191-608-20
: Sequence 20, Application US/09191608
: Patent No. 6242216
: GENERAL INFORMATION:
: APPLICANT: Lynch, Kevin J.
: APPLICANT: Burgard, Edward C.
: APPLICANT: Metzger, Randy E.
: APPLICANT: Niforatos, Wende
: APPLICANT: Touma, Edward B.
: APPLICANT: Van Biesen, T.
: TITLE OF INVENTION: Nucleic Acids Encoding a Functional
: TITLE OF INVENTION: Human Putinoceptor P2X2 and P2X4 And Methods Of Production
: FILE REFERENCE: 6194 US P1
: CURRENT APPLICATION NUMBER: US/09/191,608
: NUMBER OF SEQ ID NOS: 26
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO: 20
: LENGTH: 497
: TYPE: PRT
: ORGANISM: Homo sapiens
US-09-191-608-20

```

Query Match 3.2%; Score 93; DB 4; Length 497;
 Best Local Similarity 22.9%; Pred. No. 0.52;
 Matches 69; Conservative 36; Mismatches 86; Indels 110; Gaps 18;

```

QY 11 LLLISFVWWMIFITSONFTKLSALNISIVHYNNNSALPFPKSLI-PLKPLTER 69
   |||||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 50 LLLILYF---YMYFVIVQKSYOE-----SETPRESSITTKYKGLTTS 88
   |||||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 70 ELRIKEIIEKLDQIIPRP-----FTVNTTTSATNST-----ATLINPRDYCR 114
   |||||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 89 EHKVWDVEEYK---PREGSVFSITTRVETHTSGTGCTPESIVNHAITLSDAD--CV 143
   |||||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 115 GQDILILEVDHIGRKQKQYGDPLRA-----RMSFALTAGASG-----KYMDPNN 161
   |||||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 144 AGELDML-----GGLRTGRGVPIYQGSKTCEVFQMCPEVEDGASVSQF-L 188
   |||||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 162 GTYLVSTFLFWEQVSLSLIHPSEGASALMRAPNOGYDK-----IIFK-GK 208
   |||||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 189 GTMAPNFILIKN--SIHYKPFHFSKGNIA---DRTDGYLKRCCTPHEASDLYCPIFLGP 243
   |||||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 209 EVNGTSHVETE-----GCLTINSNAELCYLDDHDOEAFYC-----MKQNHPE 253
   |||||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 244 IYKKGESFTLAKHGKGVICVILIMDCDL-----DLPASECNFKYSFRRLDPKHVPAS 296
   |||||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 254 A 254
Db 297 S 297

```

```

RESULT 5
US-08-194-560-2
: Sequence 2, Application US/08194560
: Patent No. 6253062
: GENERAL INFORMATION:
: APPLICANT: Campbell, Judith L.
: APPLICANT: Budd, Martin E.
: TITLE OF INVENTION: A-Type DNA Polymerases
: NUMBER OF SEQUENCES: 9
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Fiehti, Hobbach, Test, Albritton & Herbert
: STREET: 4 Embarcadero Center, Suite 3400
: CITY: San Francisco
: STATE: California
: COUNTRY: United States

```

```

: ZIP: 94111-4187
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/194,560
: FILING DATE: 14-FEB-1994
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Trecartin, Richard F.
: REGISTRATION NUMBER: 31,801
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 781-1989
: TELEFAX: (415) 398-3249
: TELEX: 910 277299
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 582 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
US-08-194-560-2

```

Query Match 3.1%; Score 90.5; DB 4; Length 582;
 Best Local Similarity 20.6%; Pred. No. 1.3;
 Matches 52; Conservative 40; Mismatches 80; Indels 81; Gaps 12;

```

QY 197 NQGYDK-----IIRKGFVNGTSHV-----FTEGCLTINSNAELCEYLDD 236
   |||||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 359 NRGYSKCDIDLFFKPCPCNDTELAKIMETLICKIKYKDGYNHCLQTPYLE----- 411
   |||||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 237 RQDEAFYCKPQHPHCEALTYMTTRNREVSYLTDKENSJPHSKY-----GV----- 283
   |||||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 412 ---KLF-----LKRIVERPRTAKIVGIGERKKWYSSEIIRKFFMGKFSREL 456
   |||||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 284 EMKDKRKHDVTNCKNREKIETECOVGKRPVPGVTLQCKWITTCNOVOLDTIRKNGC 343
   |||||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 457 EELKEKNDDEGTLLIEEEEEEET---KLNPIQYMSLNK-DGNYCHRLDFCCCKWDEL 511
   |||||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 344 LKQKLYLLGDSLTRQWITYFPVVKTLKFPDLHETGIRKLLLDA--ER----- 392
   |||||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 512 GAGRIHY-TGSKYENRMI---RIIAQKPKLTQHOLFRNNILLESFNERRIPELLNK 566
   |||||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 393 -----HTQIQKK 400
   |||||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 567 YAEPEHRNIEWK 579

```

```

RESULT 6
US-08-589-756-1
: Sequence 1, Application US/08589756
: Patent No. 5846547
: GENERAL INFORMATION:
: APPLICANT:
: TITLE OF INVENTION: STREPTOCOCCAL C5a PEPTIDASE VACCINE
: NUMBER OF SEQUENCES: 5
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent Release #1.0, Version #1.30 (EPO)
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/589,756
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1164 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear

```

MOLECULE TYPE: protein
US-08-589-756-1

Query Match 3.0%; Score 87; DB 2; Length 1164;
Best Local Similarity 21.0%; Pred. No. 9.6;
Matches 93; Conservative 49; Mismatches 132; Indels 168; Gaps 24;

```

QY 144 SSPALTAGASGKYMDENNGTYLVSFT---LPEWGOVSLILLHPSSGASALMRANQY 200
      |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |
DB 450 SQTITFNATPKVLPSTAGSTKLSRFSSMGLADGNIKPDI---AAGODILSSANKY 505
      |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |
QY 201 DKIIKGFVNGTSH---VFTEGCL-----TLNSNAELCE-----YLLD 236
      |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |
DB 506 AKL-----SGTSMAPLAVAVIMGLQKQYOTPDPTOSELDLAKVYLSAATALD 558
      |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |
QY 237 RDOEAFYCKPQHMCEALTYMTNRNREVSYLTDKENSILPHRSVGVEMKRIKIDVTN 296
      |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |
DB 559 EDEKAYF--SPRQCGAGVD--AKKASATMYTDDKNT---SSKVLNNAVSDKEVTVTV 612
      |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |
QY 297 CNKREKIEE---TCQVGMKPPVPGY-----TLQGMK--ITFCNOVOLDTIKING 342
      |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |
DB 613 HMKSDKPHELYQATVOT--DKYDGKHPALAPKALITSMOKITIPANSKQVTIPI-- 667
      |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |
QY 343 CLKGKLIYLLGDSTLRQWYTFPVVTKLPDLHETGIFKRLLDAR-----RH 393
      |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |
DB 668 -----DISQ---FSKOLLQMKNGVLEGEVRI 692
      |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |
QY 394 TQIQMKH--SYRFTVQ-----LYSLD-----HDYPREIDRLSD----- 429
      |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |
DB 693 KQPTKEELMSIPYIGFGDGMLSALEPRLYDSKSSYHEEISDAKDQLDDGLQFY 752
      |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |
QY 430 --KNTAVITFGQHRPPRIDIFIRRAICQVKAIERLFLRSPATKVILIKTNIREHM-- 484
      |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |
DB 753 ALKNDFALT--TESNPWTIINVKEGV-----ENIDIDSSSE 788
      |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |
QY 485 -IETERRGDF-----HGTIH 498
      |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |
DB 789 ITETIFAGTFAMQDDRHYYIH 810

```

RESULT 7
PCT-US93-00031-9
Sequence 9, Application PC/TUS9300031
GENERAL INFORMATION:
APPLICANT: Osborn, Laurelee
TITLE OF INVENTION: ANTIBODIES RECOGNIZING THE FOURTH
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: Allegretti & Wilcoff, Ltd.
STREET: 10 South Wacker Drive
CITY: Chicago
STATE: IL
COUNTRY: US
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/00031
FILING DATE: 19930112
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: McNicholas, Janet M.
REGISTRATION NUMBER: 32,918
REFERENCE/DOCKET NUMBER: 92,306-A; D001 CIP PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: (312) 715-1000
TELEFAX: (312) 715-1234

INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 739 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US93-00031-9

Query Match 3.0%; Score 46.5; DB 5; Length 739;
Best Local Similarity 19.6%; Pred. No. 5;
Matches 88; Conservative 73; Mismatches 170; Indels 119; Gaps 2;

```

QY 11 LLLIISTVYTMHITISQNF-----TKMSALNLSIVHWNNASRSLF----- 76
      |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |
DB 5 MVLILASNLIIMKPMASQAEKIEETPESRYLAQIGDSVLTSTIDTESFFWKGQID 144
      |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |
QY 56 -----PKSLILPKPL-----TELEIKELIEMKLVITHTFTVH 94
      |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |
DB 65 SPLNGKVTNEGTTSTLMNPVSFGNHSYLTATCESKLEKGLVEYVFKPDETHLS 124
      |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |
QY 94 TTTSATHSTATILNPROTYCRGQDLILLEVRDHLGKQKQGNPFLAAMSHALIMAS 164
      |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |
DB 125 GPLEACKPITVKGSAVDY--PDRLEIDLKGDHLKSKQFLEDAKRSLEIKSEVTVI 194
      |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |
QY 154 GKVMDFNNGTYLVSFTLFW-----EGQVSS-----LLIHS----- 234
      |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |
DB 184 PVIEDI--GKVLVCRAKLHIDEMDSVTVKQAVKELOVYISPRNVIVISNPSIKLEGSS 241
      |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |
QY 191 ALMRANQY--DKIIFKGFVNGT--SHVFTEGCLT-----NSNAELCEYL----- 244
      |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |
DB 242 VYMTCSSEGLPAPELFSKRLKDNGLHLSGNATLTIAMRMEDSGIVVCGVMYICKRN 301
      |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |
QY 237 RDOEAFYCKP-----OHMPCALTYMTNRNREVSYLTIK-- 271
      |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |
DB 302 KEVELIVQEKPFVEISPGPIAAQIGSVLWTCVSGCTSPSF--SMKVLQDLSLTKVN-- 364
      |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |
QY 272 ---ENSLFHSKGVGEMKDKKHIDVTCNKRREIETCVGNH-----KRPVP--GAYTLC 322
      |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |
DB 361 SEGINSTLTLSPVSEF--NEHSYLTCTVTC--HKKLEKGIQVELYSFPRDEIEMSGLVN 417
      |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |
QY 323 GKMITFC---NOVOLDTIKINGCLKGLI 349
      |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |  :  |
DB 418 GSAVIVSCKPSVYPLDRLEIE-LIKCETI 446

```

RESULT 8
US-08-755-587-186
Sequence 186, Application US/08755587
Patent No. 6045997
GENERAL INFORMATION:
APPLICANT: Futreal, Phillip A
APPLICANT: Wooster, Richard F
APPLICANT: Ashworth, Alan
TITLE OF INVENTION: Materials and methods relating to the
TITLE OF INVENTION: Identification and sequencing of the h-ras gene
NUMBER OF SEQUENCES: 222
CORRESPONDENCE ADDRESS:
ADDRESSEE: Bell Seltzer Park & Gibson
STREET: 310 UCB Plaza, 3605 Glenwood Avenue, Pk Drawer 31107
CITY: Raleigh
STATE: NC
COUNTRY: USA
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25 (EP3)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/755,587
FILING DATE: 25-NOV-1996

RESULT 11
PCT-US94-00198-4
Sequence 4, Application PC/TUS9400198
GENERAL INFORMATION:
APPLICANT: Schering Corp.
TITLE OF INVENTION: RAS Associated GAP Proteins
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Schering Corp.
STREET: 1 Girard Farms
CITY: Madison
STATE: New Jersey
COUNTRY: USA
ZIP: 94304-1104
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: Macintosh
OPERATING SYSTEM: 6.0.8
SOFTWARE: Microsoft Word 5.1a
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/00198
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/004,824
FILING DATE: 15-JAN-1993
ATTORNEY/AGENT INFORMATION:
NAME: Lunn, Paul G.
REGISTRATION NUMBER: 32,743
REFERENCE/DOCKET NUMBER: DX0352 PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: (201)822-7255
TELEFAX: (201)822-7039
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 3079 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Saccharomyces cerevisiae
PCT-US94-00198-4

Query Match 2.98; Score 85.5; DB 5; Length 3079;
Best Local Similarity 19.3%; Pred. No. 72;
Matches 93; Conservative 85; Mismatches 160; Indels 143; Gaps 27;
QY 3 SNTLQKTLILDISFVYTMIFISQNTKLMASLNL--SISVHYMNSAKSLFPKTSL 60
DB 2515 ANNIOKILAKIRSF-----LPSLKEAMTQSMELILVKISIH-----VFETSL 2561
QY 61 IPLKPLTELKIKIIEKIDQIIPRPYVNTTSTATSTIILN--PROYCRGQO 117
DB 2562 LVOMKPEILFIVSLIDVGPREL--RSSLA-QLLMNVCHSLA--INSALPD--HRNN 2613
QY 118 LDILLEVRDHLGQRKYGDFLARNSS--PALTAGASGVNDF--NGTYLVSFTLFE 173
DB 2614 LDEISDIFAHQVKMFCFSEDDGRILQIFSSASKFNLDFINILLMYSSTYE 2673
QY 174 GOVSLSLLLIHPSGASALWRANOGYKILFKGAFNGSHVFTEGGLTNSNAELCEY 213
DB 2674 ANV-----MKTR--YKKYVLESYF-----TNS-----F 2695
QY 234 LDRODEAFYCKKPOHNP---CEALTYMTNRNREVSYLETKE---NSLFHRSKV---- 281
DB 2696 LSRASITMIVIGMSYITTEGLKAMLIETMVAIEPKITDEHFLAISHIFTYSKIVEGL 2755
QY 282 --GVEMMK-----DRKH-----IDVTNCKNR-----EKIEECQGMKPPV 315
DB 2756 DPNLIDMKHLFWSTLFLSRHPIIIEGALLFVSGNCIRLRYMAQFENSETSLIST----- 2811

QY 316 PGSYTLGG-KWITTFQNOVDITKINCCKGK-----LYVLSDILKQWYFFKVV 408
DB 2812 ----LKGRFAHFTLSKIE-----NLGIVNMEDVFTLILFIINGLSNPFI----- 2604
QY 369 KILKFDLHETGIFKKHLLDABRHQIQMKKHSIPYVTOYLSLDRYVFEHFKAS 428
DB 2857 KSTAF-----DFLKMFRNSYFEHQINQKSDHYLCYMLLYFVNLN-----SFEELGG 490
QY 429 D 429
DB 2906 D 2906
RESULT 12
US-08-323-170B-2
Sequence 2, Application US/08323170B
Patent No. 5733772
GENERAL INFORMATION:
APPLICANT: Williamson, Kim C.
APPLICANT: Kaslow, David C.
TITLE OF INVENTION: Cloning and Expression of Plasmidum
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, 8th Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.10
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/323,170B
FILING DATE: 13-OCT-1994
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/010,409
FILING DATE: 29-JAN-1993
ATTORNEY/AGENT INFORMATION:
NAME: Quine, Jonathan A.
REGISTRATION NUMBER: P-41,261
REFERENCE/DOCKET NUMBER: 015280-113100US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 315 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-323-170B-2

Query Match 2.98; Score 85.5; DB 1; Length 315;
Best Local Similarity 19.9%; Pred. No. 74;
Matches 106; Conservative 75; Mismatches 168; Indels 185; Gaps 27
QY 27 ISQNTKLMASLNLISVHYMNSAKSLFPKTSLIPKPLTEFLPIKEIIEKLCVDP 46
DB 943 ICRKHLKEFDFTLKCNVNTQYPNIEIFPPT-----LKKRKEVLKIDLDIY 990
QY 87 RPT---HVNNTSATSTATILNPRDYCGDGLDILLEVRHLSQKQVGR----- 137
DB 991 QMSKFFKFNTP-----QAKKYLNLXPY-----LIFPNHIGKELKNNFTTKNR 1036
QY 138 -----PLRRKSSPALTAGASGVNDF-----NNYYVVSF 164
DB 1037 DVKYEQSSVLSPLSSADSLGKLVLDTOETVCLTEKIYVNLNINELGSDNNTSVTF 1096

```

OY 169 TL-----FM-----EGVSLSLILHSEGSALMRARNOGYDKITIFKCF 209
DB 1097 OVHPYIDIKPEPFYPMFGCNKNKGDNIGIVELLISKOERT--GCGFHESKIDYENEN 1153
OY 210 VNCTSHVTEGCLTLNLSNAELCEYLDRODAFYMCKPQHPMCE-----ALTYMTTN 262
DB 1154 ISSDTH---ECTLHAYEN-----DIIGFNGLETTH-PREVEVEVEDAEIYLDPEN 1199
OY 463 -----REVSYLTDKENSLEFHSKYGVEAMMDRKHDIVTNCNKKREKIEETQVCMK 313
DB 1200 CENNYKGLNSVDITTLKNA-----QTYNINNKTPPELIPYNYLDELVEISQCTIKO 1255
OY 314 PVPGGYTLOGKMTTFPCNOYQDITKINGCLKGLIYLDGSTLRQWY-----YFPR 366
DB 1256 VYK-----KIKVILIT-----KNDTV---VQSESTLDDKIKYCEHENFINPR 1298
OY 467 VYKTL-----KFFD-----LHETGIEFK-HLLIDAEHTQI-QMKKHSY 404
DB 1299 VNKTEDENVEYTCNIKIEFNFIQIFCPAMDLSIVANIOMYDIYVKPFRVPOFRK---- 1354
OY 405 FYTFQLYSLIDHDYIPREIDRLSGDKNTAIYITFGQ-----HFSPFDIFIR 452
DB 1455 FNNELHKLINSEMLH-----KITKEMILYNEKXVLLHFPYVF-LPIYIK 1399

RESULT 13
US-08-818-514-6
; Sequence 6, Application US/08018514
; Patent No. 5837838
; GENERAL INFORMATION:
; APPLICANT: Reed, John C.
; TITLE OF INVENTION: BAX Inhibitor Proteins
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/818,514
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 2446
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 450 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; US-08-818-514-6

Query Match 2.9%; Score 85; DB 2; Length 450;
Best Local Similarity 18.9%; Pred. No. 3.1;
Matches 76; Conservative 54; Mismatches 133; Indels 140; Gaps 10;
OY 46 YNNNSAKSL-----FPKTSILPLKPLTELELAKIKELIKLDDQIIPRPTHYNTT 95
DB 61 WNASSKTICPECKREKWEKGFPRKVSIL-----LRDAIEKLPLPAIRLRFEDIQN 109

```

```

OY 96 TSATHTATITINPRDIYCRGDQDILLFVRDHLGQ-RKQYGDSEFLARMSSPALTAGASG 154
DB 110 NDIVQSLAFAFK-----YGNQDIPL---APNIGRANQOMGGGFF-----SGVLTG---- 151
OY 155 KYMDFNNGTGYVSTFLTFMEQGVSLSLILHSEGSALMRARNOGYDKITIFKCFPNGTS 214
DB 152 -----LTGAVAYVLLVHMYSSRESHDLVHK---ANAKMTA-----EEVYIMLEQCPMA 198
OY 215 HVTEGCLTLNLSNAELCEYLDRODAFYMCKPQHPMCEALTYMTTNREXSYLTCKENS 274
DB 199 SLXRE-----RFLSER-----VNCRLLLLTBEF--- 222
OY 275 LFHRSKYGVEAMMDRKHDIVTNCNKKREKIEETQVCMKPVPGGYTLOGKMTTFPCNOVO 334
DB 223 -FSTKYTYTLENSHRRALLM-----ELEVKALGSYKPP-----QNLN----- 258
OY 335 LDTIKINGCLGKLIYILGDSITLRQWYIFPKYVYKTLKFFDLHETGIFKKHL----- 386
DB 259 -EYKAVNPGRSILFLYALKSSPRLSLY-----LYLEDTDTFLFPIHTICPLQEDS 309
OY 387 -----LLDAERHTQIQMKH--SYFVTFQLYSLIDHDYI 419
DB 310 SCEDIYTKLLDKPTWKQWRERLVKYSFLYQULAEFAMDWL 352

RESULT 14
US-09-115-934A-6
; Sequence 6, Application US/09115934A
; Patent No. 6130317
; GENERAL INFORMATION:
; APPLICANT: Reed, John C.
; TITLE OF INVENTION: BAX Inhibitor Proteins
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/115,934A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/818,514
; FILING DATE: 14-MAR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 3209
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 450 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; US-09-115-934A-6

Query Match 2.9%; Score 85; DB 4; Length 450;
Best Local Similarity 18.9%; Pred. No. 3.1;
Matches 76; Conservative 54; Mismatches 133; Indels 140; Gaps 18;

```

Fri Oct 12 16:09:18 2001

us-09-729-454-2.rai

Page 8

```

0Y      46 YNNNSKSL-----EPKSLILPKLPLETELRLKEILEKLELOQUPRRPPTHVTT 95
Db      61 WMASSKRECECHREKHNEFPFVSL-----LRDAIEKLPPDRIALRPFEDION 109
OY      96 TSATSTATTILNRPDYTCAGDOLDILLEVRDLGO-KROYGDFELARRMSSPALTAGASG 154
Db      110 NDIIVSLAAOK-----YGNDDIPL---APNTGRANOOOGCGFF--SCVLTA---- 151
OY      155 KYMDFNNGTYLVSTFLTFMGQVSLLLIHPSSEGAALMRARNOGYDKITFGKRVTNCTS 214
Db      152 -----LTGVAVLVLLVYHMSRESHDLLVHK---AVAQWTA-----EYVVILLBDLGPM 198
OY      215 HYFTBCGLTNSNAELCEYLDDROEAIFYMKRKQMPCEALTYYTRNREVSYLETKENS 274
Db      199 SLIRE-----RLSER-----VNGRLTLTLEE-- 222
OY      275 LFRSKGVDEMKDRKHIDVTNCKRKREIETCOVMKRPVPQGTYLLOGKWITTFQNOVO 334
Db      223 -FSKTPYTIENSHRRAILM-----ELERYVALGVKRP-----ONLM----- 258
OY      335 LPTIKINGCKLGYLLLDGDSTLQWITYYPEPVKVYTLKFPDLHETGIFKKHL----- 366
Db      259 -EKYAVNGRSLEPLLALKSPRLSLLY-----LLPEYDTLPFIIRHICPLEDS 309
OY      387 -----LLDARNHQIQOMKKH--SIPEFYTOQLSIDHYI 419
Db      310 SGEDEVTKLIDLKEPTWKQMREFLVKTSFLPOLIAEFAMDWL 352

RESULT 15
US-08-553-436A-8
Sequence 8, Application US/08553436A
Patent No. 5866790
GENERAL INFORMATION:
APPLICANT: HESSE, Holger
APPLICANT: MULLER-ROBER, Bernd
TITLE OF INVENTION: DNA SEQUENCES AND PLASMIDS FOR THE
TITLE OF INVENTION: PREPARATION OF SUGAR BEET WITH CHANGED SUCROSE
NUMBER OF INVENTIONS: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Ostrolenk, Faber, Gerb & Soffen
STREET: 1180 Avenue of the Americas
CITY: New York
STATE: NY
COUNTRY: US
ZIP: 10036-8403
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/553/436A
FILING DATE: 17-NOV-1995
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/EP94/01671
FILING DATE: 20-MAY-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: DE P 4317596.1
FILING DATE: 24-MAY-1993
ATTORNEY/AGENT INFORMATION:
NAME: Mellman, Edward
REGISTRATION NUMBER: 24,735
REFERENCE/DOCKET NUMBER: P/951-117
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 382-0700
TELEFAX: (212) 382-0888
TELEX: 236925
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 766 amino acids

```

```

; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-553-436A-8

```

Query Match 2.9%; Score 85; DB 2; Length 766;
Best Local Similarity 19.9%; Pred. No. 7.6;
Matches 62; Conservative 56; Mismatches 131; Indels

[illegible]

Search completed: October 12, 2001, 16:03:55
Job time: 78 sec

Fri Oct 12 16:09:18 2001

us-09-729-454-2.rai

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OK protein - protein search, using sw model

Run on: October 12, 2001, 16:04:29 ; Search time 29.42 Seconds

(without alignments)
1416.298 Million cell updates/sec

Title: US-09-729-454-2

Perfect score: 2908

Sequence: 1 MSNFMLOKTLILISFSVY.....IHPDHYIGNIMFLNITIC 547

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR,68:.*
1: PIR1:.*
2: PIR2:.*
3: PIR3:.*
4: PIR4:.*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	2320	79.8	540	2	B45665	adult-specific 61.
2	119.5	4.1	860	2	T28227	ORF MSV067 probabl
3	117	4.0	792	2	T00082	hypothetical prote
4	112	3.9	938	2	A39160	transcription acti
5	104.5	3.6	1251	2	T45787	disease resistance
6	103.5	3.6	1237	2	S64385	probable membrane
7	104	3.5	801	2	T06783	ent-kaurene syntha
8	100	3.4	527	2	A81712	4-alpha-glucanotri
9	100	3.4	698	1	S00742	cytochrome-c oxida
10	99.5	3.4	400	2	S44810	F44B9.7 protein -
11	99.5	3.4	614	2	B43935	spat protein - Bac
12	98.5	3.4	368	2	S46435	sodium channel -
13	98	3.4	699	2	S29499	probable membrane
14	98	3.4	749	2	S61643	DNA topoisomerase
15	98	3.4	761	1	B64506	trehalose-phosphat
16	98	3.4	817	2	T43659	probable membrane
17	98	3.4	892	2	S57055	thyrotropin-releas
18	98	3.4	1025	2	T59311	exodeoxyribonuclea
19	98	3.4	1121	2	G64103	hypothetical prote
20	97.5	3.4	1046	2	T37053	hypothetical prote
21	97	3.3	679	2	S53831	hypothetical prote
22	97	3.3	1324	2	T01508	chemotaxis sensor
23	96.5	3.3	671	2	D72346	probable DNA helic
24	96	3.3	1063	2	T38420	disease resistance
25	96	3.3	1103	2	A85189	hypothetical prote
26	95.5	3.3	823	2	T34472	hypothetical prote
27	95.5	3.3	1352	2	G84473	ring-infected eryt
28	95	3.3	838	2	A84440	probable calcium c
29	95	3.3	1146	2	T18770	

30	95	3.3	1274	2	I40813	neurotoxin type F
31	95	3.2	1714	2	E71609	Ser/Thr protein K1
32	94.5	3.2	4981	2	T18489	hypothetical prote
33	94	3.2	476	2	A56510	dolichyl-diphospho
34	93.5	3.2	227	2	D81259	probable tonb tran
35	93	3.2	819	2	T07319	hypothetical prote
36	93	3.2	892	2	B36849	hypotheical prote
37	93	3.2	882	2	T28552	faty-acid synthas
38	93	3.2	1857	1	S01787	probable resistanc
39	93	3.2	2467	1	D71437	hemolysin tyc hom
40	92.5	3.2	424	1	B64216	two-component sens
41	92.5	3.2	502	2	H83716	dystonia isoform 2
42	92.5	3.2	686	2	I49298	vacuolar carboxype
43	92.5	3.2	1049	2	S67613	genome polypotein
44	92.5	3.2	2875	1	RYUTW	major core protein
45	92	3.2	891	1	KMY23W	

ALIGNMENTS

RESULT 1

B45665

adult-specific 61.9K brush border protein precursor - rabbit

C:Species: Oryctolagus cuniculus (domestic rabbit)

C:Date: 03-May-1994 *sequence_revision 03-May-1994 *text_change 04-Mar-2000

C:Accession: B45665

R:Boill, W.; Schmidt-Chanda, T.; Semenza, G.; Mantel, N.

J. Biol. Chem. 268, 12901-12911, 1993

A:Title: Messenger RNAs expressed in intestine of adult but not baby rabbits. Isolat

A:Reference number: A45665; MID:93286138

A:Accession: B45665

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1540 <BOU>

A:Cross-references: GB:212840; NID:91761; PIDN:CAA78302.1; PID:91762

C:Superfamily: rabbit adult-specific 61.9K brush border protein

C:Keywords: intestine

Query Match	79.8%	Score 2320	DB 2	Length 540:
Best Local Similarity	78.8%	Pred No. 1.9e-176:		
Matches 427:	Conservative 59:	Mismatches 54:	Indels 2:	Gaps 2:
OY	6	MLOKTLILISFSVYTWNIPTISQNTKLMSALNISVHYMNSAKSLFPTSLIFLKP	65	
DB	1	MLRKYIKLILCLALIC-VLCIISSNSTKIMGALKIPNSHYNTSMISSIPMSVSVKS	59	
OY	66	LTEFLRIKEIIEIKIDQIPRPFTVNTTTSATSTATTINPRDYCRGDDLLLEVR	125	
DB	60	LTELRLVKEILEKIDRLIPRPFTVNTTTSATSTATTINPRDYCRGDDLLLEVR	119	
OY	126	DHLGOKRGQGFPLRANSSPALTACASGVMDFNNGTYIVSTLFWESQVLSLLIHP	185	
DB	120	DYLGOKRGQGFPLRANSSPALTACASGVMDFNNGTYIVSTLFWESQVLSLLIHP	179	
OY	186	SEGASALMRARNOGYDKIFKGFVNGTSHVTECGTLTNSMLECYLDDRDQAFYCM	245	
DB	180	SEGASALMRARNOGYDRIIFKGFVNGTSHVTECGTLTNSMLECYLDDRDQAFYCM	239	
OY	246	KPOMPCCLATWTTNRREVSYLTDKNSLFRSKVGVEMAKRKIIDVTGNKREITEE	305	
DB	240	KPOMPCCLATWTTNRREVSYLTDKNSLFRSKVGVEMAKRKIIDVTGNKREITEE	298	
OY	306	TCGVGKPPVPGGYTLGGKWTTCFNOVOLDTIKINGCLGKLIYLGDSLRQMIYFFP	365	
DB	299	KCOIGKIPVPGGYTLGGKWTTCFNOVOLDTIKINGCLGKLIYLGDSLRQMIYFFP	358	
OY	366	KVVTTLKFPDLNETGIFKKHLLDAPRHQIOAKHSKSPVPTQSLSDHDYIPRPIED	425	
DB	359	KVMTLKFDFDLNETGIFKKHLLDAPRHQIOAKHSKSPVPTQSLSDHDYIPRPIED	418	

```

Query Match      4.18; Score 119.5; Db 2; Length 860;
Beat Local Similarity 18.38; Pred. No. 0.29;
Matches 132; Conservative 96; Mismatches 207; Indels 285; Gaps 31;

OY 4 NTHL---QXTLLILISFSVNTMIFLISONTF---KLSMAL-----NLISVHYWNNSAK 52
      | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
DB 126 NNMIIPEKKLIMSNTVTHILISISYIFLNGNVAVEKFKLGTGAODNNKKMIGOTLNN---182
OY 53 SLFPKSLIPLKPLETELKKEIIEKLDQOQPRPETH-----VN-----TTSA 98
      | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
DB 183 -----VLTQKQTSYIEI-EILNKLKSKEPEFRDITNNEFIYSINNLNNEALSKRN1 234
OY 99 THSTAILNPBDYTCRGDOILILEYRHLQORNOYGDGLFRRASSPALTAGASGVMD 158
      | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
DB 235 SPSSVKTEMLPBDLSL-----YLDNKSRY-----FLTSKIDGEVQ 268
OY 159 FNNQYLVLSFLFEWQSVLSLLLIHPSEGASALMRANOGDKIIIFKGRVNTSH---215
      | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
DB 269 FTVKNGICDILV-----NFIKNKQTNIPNHYW 298
OY 216 -----VTECGLT-----LNSNMLCEYLDPRDCAFYCKMPOHMPG 252
      | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
DB 299 KQGEFKKIDVAKTIFPVEFTLSTDSKQYNNRREMKRFPYNDINKKKEPDITFKAKPI 358
OY 253 -----EALYVMTTRNREVS---YLTQKENSLEFRSKV-----281
      | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
DB 359 IPIEKDIYNEALIAVYSLDEKITDGVILLDESNIDYKILONTVDIASLNTHKSPNKI 418
OY 282 ---GVEW---MKDRHIDVTCNKRKEIEEDCYGAKMPPFGSYTLLOGMWTTCNOV 333
      | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
DB 419 HKDGIMYFNINYNDEKN-----ITELKVEK-----SNDLITYNDI 457
OY 334 QUDTIRKINGCLKOKRLLYLGDSITKRWMTYFFPVV-----KTLKFFDIJHETGFKKHL 388
      | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
DB 458 NMLKFKNIHNRYSKNLIVAPLCCLIVEYSFLKSIXIVNIRLDKTYFKFSNVTYGSKL-VIL 516
OY 389 DAERHTIOJOMKKSYSYFVYFOLYSIDHDIYPREIDRSLGSKONTAIY-----IT 437
      | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
DB 517 DS-----KTHHEKYPVLDYVG-IDYLTKMKKINELNYDCKKMLINOWAVEYKEPEPE 567

```

```

Query Match      4.00; Score 117; Db 2; Length 792;
Best Local Similarity 28.58; Pred. No. 0.4;
Matches 41; Conservative 23; Mismatches 50; Indels 3; Gaps 6

QY 67 TETELURI--KEILEKL---DOQLPPRP-----FTHVNTTSA 96
      |||||  |||  |||
Db 314 TETEVLTVKKQSEKLNELADQDFPLHPRENQDLFVETPHGLKSKSHNLTLITLITNAVA 177
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

QY 99 TISTNTLLINRDYTCRGDQDLDILEVROHLQCFQWQSGDFLPAFNSFALTIAQSKVMD 176
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 374 SEVTAATGEGLROIIL-GQPMSPVITTKDKDELCKTGNALTAELSTIDQSNV-ADQELIIL 411
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

QY 159 FNGNGYLVSEFTLEFWEGQVYSILL 182
      ||||  ||:  ||:  |||||
Db 432 KKNQGIYEFELTVQKGGDTLSLRL 455
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

```

RESULT 4
A39160
transcription activator of levansase operon levK - *Bacillus subtilis*
C.Species: *Bacillus subtilis*
C.Date: 24-Jan-1992 #sequence_revision 24-Jan-1992 #text_change 02-Feb-2001
C.Accession: A39160; A69651
R:Debricouville, M.; Martin-Vershaete, I.; Klier, A.; Rapoport, G.
Proc. Natl. Acad. Sci. U.S.A. 88, 2212-2216, 1991
A>Title: The transcriptional regulator levK of *Bacillus subtilis* has domains homologous to
A.Reference number: A39160; PMID:91172782
A.Accession: A39160
A.Status: preliminary

Db 699 EV---SSDSQSLSTIMVYLQNLKVLDSLQCELEDEDIQIPKRLKRLGTAIKELPSLMH 755
 QY 473 ----VIKTENIREMH 484
 Db 756 LSELVYLDELQNCRLH 771

RESULT 6

64385
 Probable membrane protein YGR090W - yeast (*Saccharomyces cerevisiae*)
 N:Alternate names: hypothetical protein G4639
 C:Species: *Saccharomyces cerevisiae*
 C>Date: 17-May-1996 #sequence_revision 17-May-1996 #text_change 20-Jun-2000
 C:Accession: S64385; S64395
 R:Medler, H.; Scharfe, M.; Medler, E.; Wandutt, R.
 Submitted to the Protein Sequence Database, May 1996
 A:Reference number: S64356
 A:Accession: S64385
 A:Molecule type: DNA
 A:Residues: 1-1237 <MED>
 A:Cross-references: EMBL:Z72675; NID:q1323132; PID:q1323133; MIPS:YGR090W
 A:Experimental source: strain S286C
 R:Hernandez, K.; Weber, N.; Wipfl, P.; Schmidheini, T.
 Submitted to the Protein Sequence Database, May 1996
 A:Reference number: S64392
 A:Accession: S64395
 A:Molecule type: DNA
 A:Residues: 1-1237 <HER>
 A:Cross-references: EMBL:Z72675; NID:q1323132; PID:q1323133; MIPS:YGR090W
 A:Experimental source: strain S286C
 C:Genetics:
 A:Map position: 7R
 C:Superfamily: *Saccharomyces cerevisiae* probable membrane protein YGR090W
 C:Keywords: transmembrane protein
 F:385-401/Domain: transmembrane #status predicted <TM>

Query Match

Best Local Similarity 3.6%; Score 103.5; DB 2; Length 1237;
 Matches 102; Conservative 87; Mismatches 175; Indels 165; Gaps 28;

QY 30 NFKTLMALNLSIVHYMNSAKSLFPRKSLPLAKLETETELRIKEITELKLDQOIPRPF 89
 Db 284 NFKTRISINLIGFPKVFEPKLLPNRNCI-----RIAQ--EKEKESLPAPPL 331
 QY 90 THVNTTSATNSTATILNPRDTYCRGQDLDILLEVDHLGQRKQYGDPLRARM----- 143
 Db 332 YNFVSLSSSTH-----ENYLK-----YLTKKQTESFEVATVIGRLML 370
 QY 144 -----SSPALTAGAGKVDNFNGTLYVSFTLFE-----GVLSLLLIHSEGASALM 193
 Db 371 QORGFSSMSHSGSLG-----GQTF--EFTILMAALLNGGGINSKNLIH----- 414
 QY 194 RARNQGTDKI--IFKG--KFNQGTSHVTECGTLNSNAE-----LCEYLDPRQDEAYC 244
 Db 415 -----GFSYQLFKGVYIKYL--ATMDLCHDGLQFHSNENSSSPASKYIDEGFOTPLF 468
 QY 245 NKPOHMPCEALTYMTNRNFEV--SYLTDKENSLEHRSKVGVEKKOR--KHIVTQCNRE 301
 Db 469 DKSTKV--NLTGKMTYSSVQILKEAGETILRLMN-----VQDDFSNITLITNISFD 519
 QY 302 KLEETCOVGKRPVPGGTYTLGGKMTTFECNQVQDLTIKINGCLKGLIYLLDSTLRQMI 361
 Db 520 NLK-----YDL-----CYDQQLPLGKYNN--LETSLAATFGSMERVKFI 556
 QY 362 YVFPKVVKTLKFPDLHETGIFKKLLDLAEHRTQIQM--KHSYFPVTFOLYS-----LI 414
 Db 557 -----TLENFLAKITINVARVALCDRIKYIQIEMVQGSQSDPLTKRKVSNNGCHNF 608
 QY 415 DHDYI-----PRELDRL--SGDKNATVITFGGHPRPFDITFIRAKIYQKAIER-- 463
 Db 609 NDEYRVKLLVNPSECDKLVTKGPAHSETMSTEAIVKFN-----WGIKSSLRRFK 659

QY 464 -----LEIUSPATKVIKTENI-----KEMHIEFERGFHCTI 497
 Db 660 DQSTIRHCCWSTSSSEPTISTIVNALQKHVSKKAOISNETIKKPHNLT 708

RESULT 7

106783
 ent-kaurene synthase A (EC 2.5.1.-) - garden pea
 C:Species: *Pisum sativum* (garden pea)
 C>Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 11-May-2000
 C:Accession: T06783
 R:Alt-All, T.; Swain, S.M.; Reid, J.B.; Sun, T.P.; Kamtva, Y.
 Plant J. 11, 443-454, 1997
 A>Title: The 1S locus of pea encodes the gibberellin biosynthesis enzyme ent-kaurene synthase
 A:Reference number: Z15810; MUID:97260958
 A:Accession: T06783
 A>Status: preliminary; translated from GB/EMBL/DDBJ
 A:Molecule type: mRNA
 A:Residues: 1-801 <AIT>
 A:Cross-references: EMBL:U63652; NID:q2160543; PID:AAH58e22.1; PID:q1160544
 A:Experimental source: cultivar Torsdag
 C:Genetics:
 A:Gene: 1S
 C:Function:
 A:Description: converts the geranylgeranyl diphosphate into copalyl di diphosphate
 A:Pathway: the first step of the gibberellin biosynthesis
 C:Keywords: transferase

Query Match

Best Local Similarity 3.5%; Score 103; DB 2; Length 801;
 Matches 124; Conservative 91; Mismatches 197; Indels 214; Gaps 4

QY 2 SSNTMLQKTLILISFVYTW-----MIFITSONPFRKIMALNLSIVHYMNSA 1
 Db 154 SAHRIILNTLACVIALR--SNMMSSEKQDGMF--FRENISKLENEHEHMPIGF----- 205
 QY 52 KSLPRKSLP-----LKL-----TETELRIKEITELKLDQIP--EPRPHVNTTSATHSI 102
 Db 206 EVAPF--SLLEGARQIKRPLMCPNOSP--LKNIFEKRDKLRIKPEIMHVPF----- 295
 QY 103 ATILNPRDTYCRGQDLDIL--LEVRDHLGQRKQYGDPLRARMSSFPALTAVASAKVYNN 141
 Db 256 -TLHSLEGSGIDMKVQLKIQSD-----GSLP-----FPRSTNALMLQIKLQNG 401
 QY 162 GTLYVSFTLFEQVSLSLLIHSEGASALM--RARNQGTIKIFKGF--VNGTSH 215
 Db 302 LKTLNNVYKKFNGVP-----NVYVVDLEFHIWVDRLERLISR--FKHEIDCNMYVSK 375
 QY 216 VFTEGQL--TLNSNAELCEYLDPRQEA-----PYCKRP 247
 Db 357 IMSEKGIQARNSNQ-----DIDDTAMAFRLRLRHGHQVSAHVRFKHFERNGEFFCPAG 410
 QY 248 OHMFC--FALTYMTNRNREVSYLTDKENSLEHRSKVGVEKKOR--KHIVTQCNRE 301
 Db 411 Q--CTQATVGMNLFRAVOVLPFGKEKILHAKHSAKAVLEKKEANELLQKMIIRKPL 467
 QY 292 -----IDVTQCNKREKTEETG--QVGKRPVPGGTYTLGGKMTTFECNQVQDLTIKING 361
 Db 468 EEVYALADRWYANLDRIETRFYIDYGASESVMIQKTLIRMAVNNVNNVLLARLQNN 527
 QY 343 CLKGLIYLLDSTLRQMIYVFPKVVKTLKFPD--LHETGIFKKLLDLAEHRTQIQMKH 401
 Db 528 CQAGHLL-----EM-----NVQIOTYLESRLGEGSLSKRDLI----- 560
 QY 402 SYRVVFPOLYSLDHDYIPREIDRLSGDKNATVITFGGHPRPFDITFIRAIQVQKAI 461
 Db 561 AVPLATGSIFF-----PERSHRLMAATTAALLT----- 591
 QY 462 ERLFSPATKVIKTENIRBMHIEFERGFHCTIHYLLMKDIFPD-----LN 510
 Db 592 -----KCYVNEDLRKDF--AKKFNHDIDVDYSLIARBMKMKNTKEHLEIVSLSLAT 619

QY 511 VGIIDAMDTIAYGTDIHPDDHVG 536
 DB 640 IGEI-SMDVRLSYG-----HEIG 656

RESULT 8
 A81712

4-alpha-glucanotransferase TCO462 [Imported] - Chlamydia muridarum (strain N199)

C:Species: Chlamydia muridarum, Chlamydia trachomatis MOPn

C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 02-Jun-2000

C:Accession: A81712

R:Read: T.D.: Brundham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey, C.; Dodson, R.; Gwinn, M.; Nelson, M.; Deboy, R.; Kolonay, J.; McClarty, G.; Salzberg, M.; et al. 2000

C:Title: Genome sequences of Chlamydia trachomatis MOPn and Chlamydia pneumoniae AR39.

A:Reference number: A81500; MUID:20150255

A:Accession: A81712

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-527 <TFT>

A:Cross-references: GB:AE002303; GB:AE002160; NID:g7190392; PID:AAF39223.1; PID:g719040

A:Experimental source: strain N199 (MOPn)

C:Genetics:

A:Gene: TCO462

C:Superfamily: 4-alpha-glucanotransferase

Query Match 3.4%; Score 100; DB 2; Length 527;
 Best Local Similarity 20.0%; Pred. No. 5.1;
 Matches 92; Conservative 68; Mismatches 158; Indels 142; Gaps 25;

QY 65 PITELELR--KEILEKIDQOI--PPR-----PETHVNTTSAHSTATILNPRDTCR- 114

DB 2 PLRSRLRIQNSPIKVMNOVDTSFKHICVPLFSHTNCSGIGEFILPMIDCTL 61

QY 115 -GQDDILEVRDHLGORKOYGDPLRAMSSPALTAGASGKWMDFNNGTLYVSFTLFE 173

DB 62 GGFQILDILPINDTSCSSPPNS-----ISLISL-----NPLHLISALPYK 103

QY 174 GOVSLSLILIHPSGASALWRARNOGYDKII-FKGFVNQTSVHTFEC-GTLNSNAELC 231

DB 104 EEVSSSRKLIQEMQRLSLSQV---NKEKVPKRAFF---KEKFPVCKSKNLNHHDFC 157

QY 232 EYLDKDDQ-----EAFYCKMPQMPCEALITYTRNREVSYLTKQNSL-----FHRS 279

DB 158 DFC-EKEKYLWLRPYALFCSIREHLNLPINHWSTYDLSYISQHEHTFADIDFYSYLQ 216

QY 280 KVGVEKMKD-RKHIDVTNCKNREKI-----EETCOV-----GKMPVPGGYTLQ 322

DB 217 YLGFEDMQVQRKADHKGCLIKGDIPLISKSDCVAFYKRYSSSESYSQSP-PDFYNAE 275

QY 323 GK-WITFCNOVOLDITKINGCLKGLIYLLGDSLBMQIYYFPKVVTKL----PFDL- 376

DB 276 GQMMNLPITYM-----FTLQADAYHMKW--EKLRYANPFSLY 311

QY 377 --HETGIFKRLDLDAERHTQLOMKKHSYFVYFQULSLIDH-----DYIPEID 424

DB 312 RLDRVGLFR-----FWWDELGRGREPODDPKYLDLGGTD 347

QY 425 RLSCOKNTAIVITFGOHFRFPIDIFIRRAIGYOKAIELR 464

DB 348 ILSHLKASSMLPIGEDLGTIPV-----VKQALESL 379

RESULT 9

S00742

Cytochrome-c oxidase (EC 1.9.3.1) chain I - Tetrahymena pyriformis mitochondrion

C:Species: mitochondrion Tetrahymena pyriformis

C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 07-Dec-1999

C:Accession: S00742

R:Zlatar, Z.; Suyama, Y.

C:Gene: 12; 357-368, 1987

A:Title: The cytochrome oxidase subunit I gene of Tetrahymena: a 57 amino acid NH2-termi

A:Reference number: S00742; MUID:88184706

A:Accession: S00742

A:Molecule type: DNA

A:Residues: 1-698 <ZIA>

A:Cross-references: EXBL:X06133

C:Genetics:

A:Genome: mitochondrion

A:Genetic code: SGC6

A:Start codon: ATA

C:Superfamily: cytochrome-c oxidase chain I; cytochrome-c oxidase chain I homology

C:Keywords: chromoprotein; copper; electron transfer; heme; iron; magnesium; membrane

F:59-617/Domain: cytochrome-c oxidase chain I homology <COI>

F:111,538/Binding site: heme a iron (His) (axial ligand) *status predicted

F:401,450/451/Binding site: copper (His) *status predicted

F:401-405/Cross-link: 1'-histidyl-3'-tyrosine (His-Tyr) *status predicted

F:405/Binding site: oxygen (Tyr) *status predicted

F:528/Binding site: magnesium (His) (shared with chain II) *status predicted

F:536/Binding site: heme a3 iron (His) (axial ligand) *status predicted

Query Match 3.4%; Score 100; DB 1; Length 698;
 Best Local Similarity 19.2%; Pred. No. 7.5;
 Matches 105; Conservative 81; Mismatches 194; Indels 166; Gaps 30;

QY 23 MIFISQNTKLM-----SALNLSIVH---YWNNAKSLPEKTSLLPLKPLTELELRK 74

DB 1 MLTYFN-SFDNNWDFEIQTSKPKSVNMYNLYNKKIKLF--TYLNDLR-----K 48

QY 75 EIEKIDQOIPRPF-----THVNTTSAHSTATILNPRDTCRQDDIDLEY 124

DB 49 HILKKVYITNKRRAINLYFSWTLGSGAALMTMRBLMHPSPFFKDSRLYQVY 108

QY 125 RDHLGORKOYGDPLRAMSSPALTAGASGKWMDFNNGTLYVSFTLFE 176

DB 109 TAH-----GLINFEVYVPLTFEGFANFLIPYHVSQDAVYPLNSIGFWIPQGYI 160

QY 177 SLSL-LIHPSGASALWRARNOGYDKII-FKGFVNQTSVHTFECGTLNSNAELCEYLD 235

DB 161 LKAKIFELPQ-----FWRV---YKTSFSPLEKMKY-----NOKRYKYN 199

QY 236 D-----KQCAFY-----CKMPQH-----MPEALITYTRNREVSYLTKQ 272

DB 200 DYLFYDLFLKEITDDHSFFMKARKYIKLPQVSVSPVPLKLMKMTMINPESF----- 254

QY 273 NSLFHRSKVGVEKMKDKRHIDVTNCKNREKIEETCOVGMKPPVPGGYTLQGKATTFQNG 332

DB 255 --WYASRY---VQSRKKVYVTKCSARTL-----TTAGMT----FITPSS- 292

QY 333 VOIDTKINGCLKGLIYLLGDSLBMQIYYFPKVVTKLKFPLDHEGIFKRLDLDAER 392

DB 293 ---NRTYTG-----VGSODILLSVFAISITISFTNLL--ITRRLAMPQMK 336

QY 393 HTOIQMKHSYFVYFQULSLIDHVIPEIDRLSGOKNAIY-ITFGOHFRFPIDIFI 451

DB 337 HRRV-----LMPFYTISIFTL-----RMLATTIPVLAAGVYMAVAFRHHQ-----TTF 381

QY 452 RRAIGYOKAIEL--LFLRSAPKVIKTE---NIFEMHIEERPGDF-----G 495

DB 382 EYAYGDDPLSOLHFWFGHPEVYLLIPTGFIINLVPRNNTRVASKHMHIMAYVMA 441

QY 496 YIRHYL 501

DB 442 YMGYLV 447

RESULT 10

S44810

R4489.7 protein - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 09-Sep-1997

C:Accession: S44810

R:Antonaci, L.

A:Residues: 1-168 <RCU>
 A:Cross-references: EMBL:X72582
 R:Boistad, A.L.
 submitted to the EMBL Data Library, August 1994
 A:Reference number: S51507
 A:Accession: S51507
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-156, 'K', 158-368 <RC07>
 A:Cross-references: EMBL:X72582; NID:q530295; PIDN:CAA51172.1; PID:q530296

Query Match 3.4%; Score 98.5; DB 2; Length 368;
 Best Local Similarity 21.5%; Pred. No. 4;
 Matches 67; Conservative 35; Mismatches 132; Indels 77; Gaps 15;

```

QY 49 NLSISVHY WNNSSAKSLPRKSLPLKPLETELKIKELIKLDOQIPRPPTHTVNTTS 97
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 57 NGSDVOYRMVGNVENRTPKKE-DPASPLMGLGNVAGRL-----QTLTVNTEKQTLLEI 110
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 98 ATBSTATIINPRDYTCRGDQDLI-----LLEVRDLGGRKQYGGDFLARMS 145
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 111 RTNRYHTLMNPKDSQADDQYRVNHYFKGKLGSSKIDVTSRL-EYKONNGDAGKQAE 169
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 146 PALTAGASGVWDFNNGTYLVSPFLMEQVSLSLLIHSEGSALMRRNQ-----Y 200
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 170 SVL-----PFDADYIYSNNFFKADKFGFRL-----GYOHKMWGHSVVGQPF 212
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 201 DKLIIFKGFVNGTSHVFTECGTLTNSA-----ELCEYLDLDDQDAFY----CKMK 247
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 213 NKGTUNYFIINFESRYTLPMWGFSAELNAYNNVNNKFAIYNNKKGKQGFGELEATLY 272
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 248 QHMPCEALTYMTRRREVSITLDKENS---LFHSKGVGVEMKDKRKHIDVNCNKR-K 302
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 273 QHTP-----LYKTNNVELSF--DFEGGYPTVTHQYV-----VSANDSKYEYV 315
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 303 IEFTCGVGNKP 313
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 316 MLPTLQVSYKP 326

```

RESULT 13
 S29499
 C:Species: Rattus norvegicus (Norway rat)
 C:Date: 13-Jan-1995; #sequence_revision 13-Jan-1995; #text_change 09-Jun-2000
 C:Accession: S29499; S43503; S29715; S41158
 R:Llinquegila, E.; Voilley, N.; Waldmann, R.; Lazdunski, M.; Barbry, P.
 FHS Left, 318, 95-99, 1993

A:Title: Expression cloning of an epithelial amiloride-sensitive Na(+) channel. A new ch
 A:Reference number: S29499; MUID:93170495
 A:Molecule type: mRNA
 A:Residues: 1-699 <LIN>
 A:Cross-references: EMBL:X70521; NID:q433909; PIDN:CAA4916.1; PID:q433910
 R:Carassa, C.M.
 submitted to the EMBL Data Library, March 1994
 A:Reference number: S43503
 A:Accession: S43503

A:Molecule type: mRNA
 A:Residues: 2-598, 'DV', 601-699 <CAN>
 A:Cross-references: EMBL:X70497; NID:q458845; PIDN:CAA4905.1; PID:q458846
 R:Carassa, C.M.; Horisberger, J.D.; Rossier, B.C.
 Nature 361, 467-470, 1993
 A:Title: Epithelial sodium channel related to proteins involved in neurodegeneration.
 A:Reference number: S29715; MUID:93156815

A:Molecule type: mRNA
 A:Residues: 2-194, 'P', 196-230, 'GAA', 234, 'LPAYATTI', 243-598, 'DV', 601-699 <CAN>
 A:Cross-references: EMBL:X70497
 R:Carassa, C.M.; Schild, L.; Buell, G.; Thorens, B.; Gautschi, I.; Horisberger, J.D.; Rd
 Nature 367, 463-467, 1994
 A:Title: Amiloride-sensitive epithelial Na(+) channel is made of three homologous subun

A:Reference number: S41158; MUID:94150624
 A:Contents: annotation
 C:Superfamily: human amiloride-sensitive sodium channel protein; fibronectin type I
 C:Keywords: glycoprotein; sodium channel; transmembrane protein
 F:110-162/Domain: transmembrane #status predicted <TM1>
 F:422-457/Domain: fibronectin type I repeat homology <IPR>
 F:567-613/Domain: transmembrane #status predicted <TM2>
 F:425,539/Binding site: carbohydrate (asn) (covalent) #status predicted

Query Match 3.4%; Score 98; DB 2; Length 699;
 Best Local Similarity 19.2%; Pred. No. 11;
 Matches 89; Conservative 53; Mismatches 157; Indels 164; Gaps 18;

```

QY 11 LLLISVYVMTITISQNTKMSKLSISVYNNSSAKSLPRKSLPLKPLETE 70
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 115 VLMICTGKMYQFALLFEELYSVSLNINL-----NSDKLVPAVTCILNRYTE 168
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 71 LRKEIIEIKLD-----QIPRP 89
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 169 --IKEELIEDRIETQTLFDLYKNSSYTRQAGARRSSDILGAPPHLOKRLTPRPY 226
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 90 THVTTTSAHSTAILNPR-----DYCRGDQDLILEVRDH 127
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 227 S--GRTARSGSSVSDNNPDYDRDKKIGFOLCNOKSDCFYOTYSSG--VDAYRE---- 278
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 128 LGORQYGGDFLRAMSSPALTAGASGVWDFNNGTYLVSPFLMEQVSLSLLIHSE 187
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 279 -WYFPHYINLSRLSDTSPALEBALG-----NFTICFNPAPCONANS-KFHHPMY 330
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 188 G-----ASALRRANQYDKITFKKFLVNGTSHVFTECGTLTNSAEIYELDRD 238
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 331 GNCYTNDRKNSNLSMSSMP-----VNN-----GLSLTTEEDNFI---- 368
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 239 QEAFYCMKPOHPCFALTYMTRRREVSITLDKENSIFHSKGVGVEMKDKRKHIDVNCN 298
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 369 -----PLSTVTGARVYHGGDEPARNMDGCFNLRPVEVISMRKA--LDISLGN 418
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 299 KREKIEFTGVGMKPPVGGITLQGMHTTFGNVOYDITIKNGCLKGLIYLGDSTLR 358
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 419 YGDCETGSDVPVKNLYPSKYTOO-----VCIHSCQENMIRKCGCA--- 460
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 359 QMIYV-FPVVKTLEF-----FDLHETGIFKK 384
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 461 -YIYPRKPGVEFCYVRKQSGWCYCYKLGAFSLSDLGCFSK 502

```

RESULT 14
 S61643
 Probable membrane protein YOR081c - yeast (Saccharomyces cerevisiae)
 N:Alternate names: hypothetical protein 02964; hypothetical protein YOR2964c
 C:Species: Saccharomyces cerevisiae
 C:Date: 09-Mar-1996; #sequence_revision 12-Apr-1996; #text_change 21-Nov-1997
 C:Accession: S61643; S66966; S66964
 R:Benes, V.; Andrade, M.A.; Rechmann, S.; Teodoru, C.; Banerji, A.; Sander, C.; Vale

submitted to the EMBL Data Library, December 1995
 A:Description: Nucleotide sequence and analysis of a 130 kb fragment of yeast chrom
 A:Reference number: S61643
 A:Molecule type: DNA
 A:Residues: 1-749 <BEN>
 A:Cross-references: EMBL:X94335; NID:q1262139; PID:e217721; PID:q1164929
 R:Voss, H.; Benes, V.; Rechmann, S.; Teodoru, C.; Schwager, C.; Paces, V.; Ansoorge,

submitted to the Protein Sequence Database, July 1996
 A:Reference number: S66965
 A:Accession: S66966
 A:Molecule type: DNA
 A:Residues: 1-749 <VOS>
 A:Cross-references: EMBL:Z74989; NID:q1420242; PID:e251992; PID:q1420243; MIPS:YOR080
 A:Experimental source: strain S288C
 R:Bohn, C.; Bolotin-Fukuhara, M.; Dalgman-Fornier, B.; Dang, D.V.; Valens, M.
 submitted to the Protein Sequence Database, July 1996
 A:Reference number: S66929

A:Accession: S66964
 A:Molecule type: DNA
 A:Residues: 435-749 <BOH>
 A:Cross-References: EMBL:Z74989; MIPS:YOR081c
 A:Experimental source: strain S288C
 C:Genetics:
 A:Map position: 15R
 C:Keywords: transmembrane protein
 F:211-227/Domain: transmembrane status predicted <TM1>
 F:327-343/Domain: transmembrane status predicted <TM2>

Query Match 3.48; Score 98; DB 2; Length 749;
 Best Local Similarity 20.48; Pred. No. 12;
 Matches 120; Conservative 70; Mismatches 224; Indels 174; Gaps 27;

```

OY 5 TMLCKTLILISFVYVWMIPIISQNFKLMSALMSIVHWYVNSAKSLFPTSLPLK 64
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 37 TLSRKQLLI-----SDLSQKKHAIISYQVNDIASRLDLDTGLSEBK 78

OY 65 PLTEELRIKEIIEKLDQIIPRPFTVNTTSATHSATILNPRTCYRGQDLILEV 124
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 79 TIDESSLYNKLLDPLTRM-----KHLRTTHD-YHRLIYLITKVRNLMNNVNLIR 132

OY 125 RDHLGCKQYGGDFLRANMSSPALTAGASKYVDFNNGTYLVSFTLFWEGVSLSLI 184
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 133 HSHTG-TKQIHIDYLEE--SQAVLTA-----LIH 158

OY 185 PSE-----GASALMRANQGYDKIIFKGFVNGTSHVTEGGLTNSAELCEYLDR 237
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 159 QSNMNHVYLLGLQOTR-RNIGRTALVLSGSTFLFHLIGVLAAL--FESLMKPYISG 214

OY 238 DQ-----EAFYCKMPQHPCALTYMT-TRNREYSLDKENSLFHRKSVGEYMKDKRK 291
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 215 SSAGATVASFVY--HTQETPSLLTNLMNEFNFD-DNS-----253

OY 292 IDVTNCKREKIEETQYGMKPPVPGYTLQCKWITFCQVQVOLDTIKINGLKLITL 351
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 254 -KSPNENLIKISRQNG-----TMMNQPLIN-----MLSF 286

OY 352 LGDSTLRQ-----WITYPKVYKTLKFFDL-----HETGIFKKH 385
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 287 LGNLTREAVNKTGKILNITVSPASIVEQPKLLNLTAPVILMSAVCASGLPGVPST 346

OY 386 LLDAERHTQIQMKKHSYPTFQYSL-----IDHDYIPREIDRLSGDKNTAVITFG 439
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 347 PLFEKDPHT---GKIKMGATNLHLSNMKMDGSDND--APISKLSMFRNVDHITACQ 400

OY 440 QHRRPPIDIFIRALIG--VOKAIEKFLIRSPATKV--IKTENIREMHIEITERFGDFHG 495
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 401 VNIHVEPLKFSNCTVGEIEKEITARF-RNOYTKIKFPSDETIFHFDILKEL--EFHP 457

OY 496 YIHLYIMKDIKDLNVGIIIDAMDITAVGTDTHPPDHVIGNQIMFL 543
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 458 YL-MTKLKLHFLQGYSG---NVTILPLDSMVGQFHEVLAKNSQLFL 499

```

RESULT 15

B64506
 DNA topoisomerase (EC 5.99.1.2) - Methanococcus jannaschii

N:Alternate names: topoisomerase I
 C:Species: Methanococcus jannaschii

C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000
 C:Accession: B64506

R:Butt, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake,
 Iron, J.D.; Sadow, P.M.; Hanna, M.C.; Kirtress, E.F.; Melnick, K.G.; Merrick, J.M.; Glodok, A.;
 Science 273, 1058-1073, 1996

A:Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese, C.
 A:Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii
 A:Reference number: A64300; MUID:96337999

A:Accession: B64506
 A:status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA
 A:Residues: 1761 <BOU>
 A:Cross-References: GB:U67605; GB:U77117; NID:q1592227; PILEN:AA89673.1; PID:q1592227
 C:Genetics:
 A:Map position: FOR1633497-1635782
 A:Start codon: GTG
 C:Superfamily: DNA topoisomerase I topa
 C:Keywords: Isomerase

Query Match 3.48; Score 98; DB 1; Length 761;
 Best Local Similarity 21.58; Pred. No. 12;
 Matches 53; Conservative 40; Mismatches 76; Indels 78; Gaps 13;

```

OY 253 EALTMTNREVSYLTDK--ENSLFHSKGVEMM-KDKKH-LYTGNCKEKI-----304
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 288 EAVSYRKISPKETQELIAQKLYENALLSYRTSSQKLFKDKRYLEDILNLIKHHVYCKMA 347

OY 304 EETQYGMKPPVPGSYTLQK-----WITFC 340
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 348 ERLKENLKP-----VEGKKEDPAHPAIIHVIDIPKEELSEKEEIVDLIARRILAPW 400

OY 331 NOVQDLTIKINGLKLILYLDGSTLRQ-W--IYPPKVVYKTLKFEULHERGTHKKHL 407
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 401 DNAREYLVNWKIDIKGEKFKLSGRTVKCGHEIYIFPK-----FDEIELPLKNDI 453

OY 388 LDAERHTQIQMKKHSYPTVQYSLIDHDYIPREIDRLS-GKNTAIVTPQGHFPPPP 446
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 454 IKVEKITIT--RKETQPKRYTVASII-----KELERGLGTRATRAEI-----405

OY 447 IDIFIRK 453
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 496 IDKLIRK 502

```

Search completed: October 12, 2001, 16:04:32
 Job time: 115 sec

Fri Oct 12 16:09:18 2001

us-09-729-454-2.rpx

Page 9


```

Db 120 DYLHOK EYGGDFLBARMFSPALKAAGSKVTDNNGVTLVSFTLEWGGVSLVLLTHP 179
QY 186 SEGASALWRANOGYDKIIFKPKFVNGSHVFTECGLTNSAELCEYIDDDKQOEFTOM 245
Db 180 SEGASALWRANOGYDKIIFKQGFNGSHVFTECGLTNSAELCEYIDDDKQOEFTOM 239
QY 246 KPOHNPCEALTYMTTNREVSLTLDKENSLEHRSKGVEMMDKRDHIVNCKREKIEE 305
Db 240 KPOHNPCEALTYMTTNREVSLTLDKENSLEHRSKGVEMMDKRDHIVNCKREKIEE 298
QY 306 TCOVGAKPPVPGGTYLQGMKITTENNOYDITKINGCLAKGLIYLGGSTLQWATYFP 365
Db 299 KCOJGMKIPVPGGTYLQGMKITTENNOYDITKINGCLAKGLIYLGGSTLQWATYFP 358
QY 366 KYVATLKEFEDLHETGTFKHHLLDAERHTQIQMKHSHPEVTFOLYSLLDHDIYPREIDR 425
Db 359 KYVATLKEFEDLHETGTFKHHLLDAERHTQIQMKHSHPEVTFOLYSLLDHDIYPREIDR 418
QY 426 TSGDKNTAIVITFGOHFRFPIDIPIRRAIGVOKATERLEFRSPATVITKTEHREHNI 485
Db 419 LIGDKDTIVITFGOHFRFPIDIPIRRAIGVOKATERLEFRSPATVITKTEHREHNI 478
QY 486 ETERFGDHCYIHYILMKDIFDLNGLIDAMDITAVGDTIHPDHVIGNOINMFLNY 545
Db 479 EAERFGDHCYIHYILMKDIFDLNGLIDAMDITAVGDTIHPDHVIGNOINMFLNY 538
QY 546 IC 547
Db 539 IC 540

RESULT 2
LEVR_BACSU STANDARD: PRT: 938 AA.
ID LEVR_BACSU P23914;
AC 01-MAR-1992 (Rel. 21, Created)
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE TRANSCRIPTIONAL REGULATORY PROTEIN LEVR.
GN LEVR.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE-91172782; PubMed-1900939;
RA Debarbouille M., Martin-Veststrate I., Klier A., Rapoport G.;
RT "The transcriptional regulator Levr of Bacillus subtilis has domains
RT homologous to both sigma 54- and phosphotransferase system-dependent
RT regulators."
RL Proc. Natl. Acad. Sci. U.S.A. 88:2212-2216(1991).
RN [2]
RP SEQUENCE FROM N.A.
RA STRAIN-168;
RX MEDLINE-97286548; PubMed-9141695;
RA Parro V., San Roman M., Galindo I., Purnelle B., Bojotin A.,
RA Sorokin A., Mellado R.P.;
RT "A 23911 bp region of the Bacillus subtilis genome comprising genes
RT located upstream and downstream of the lev operon."
RL Microbiology 143:1321-1326(1997).
CC -1- FUNCTION: INVOLVED IN POSITIVE REGULATION OF THE LEVANASE
CC OPERON WHICH COMPRISES THE LEVEFG GENES FOR A FRUCTOSE PTS
CC SYSTEM, AND SACA FOR LEVANASE.
CC -1- PTM: PHOSPHORYLATED AND INACTIVATED BY VIA THE PHOSPHOTRANSFERASE
CC SYSTEM (POSSIBLE).
CC -1- SIMILARITY: THE N-TERMINAL REGION CONTAINS A SIGMA-54 FACTOR
CC INTERACTION ATP-BINDING DOMAIN.
CC -1- SIMILARITY: IN THE C-TERMINAL HALF; BELONGS TO THE TRANSCRIPTIONAL
CC ANTERMINATOR BGLG FAMILY.
CC -----
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -

```

```

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.ebi.ac.uk/submit/
CC or send an email to license@ebi.ac.uk).
CC -----
DR EMBL: M60105; AAA2572.1; -
DR EMBL: X92868; CAA3460.1; -
DR EMBL: Z99117; CAA14650.1; -
DR PIR: A9160; A9160.
DR Subtilist; Bg1677; levr.
DR InterPro; IPR001550; -
DR Pfam; PF00874; Bgic_antitermin; 1.
DR PROSITE; PS00654; ANTERMINATORS_BGLG; 1.
DR PROSITE; PS00675; SIGMA54_INTERACT_1; FALSE_NEG.
DR PROSITE; PS00676; SIGMA54_INTERACT_2; 1.
DR PROSITE; PS00688; SIGMA54_INTERACT_3; FALSE_NEG.
DR PROSITE; PS00945; SIGMA54_INTERACT_4; 1.
DR Transcription regulation; Activator; DNA-binding; ATP-binding;
KW phosphorylation.
FT DOMAIN 120 351 SIGMA-54 FACTOR INTERACTION (POTENTIAL).
FT NP_BIND 148 155 ATP (POTENTIAL).
FT NP_BIND 218 227 ATP (POTENTIAL).
FT SIMILAR 411 572 TO THE N-TERMINI OF BGLG FAMILY.
SQ SEQUENCE 938 AA; 106186 MW; 167025A2AEB5696 CRC64.

```

```

Query Match 3.94; Score 112; DB 1; Length 948.
Best Local Similarity 21.0%; Pred. No. 0.39;
Matches 130; Conservative 100; Mismatches 213; Indels 176; Gaps 46;

```

```

QY 3 SNTMLOKTLILISPSVYTWMIPIISONFKLMSALNLSI-SVHWNSKSLNPKFTSLI 61
DB 53 NNLVRSKVKYKITEPVRYIVYELAELEFKKHQDMEVNDLQASGNSKNDHSHSTN 112
QY 62 PLKPLTETELRIKEIIEKLDQI--PRPETHVNTTSATSHATILNRPDYCRDLD 119
DB 113 PLEDMIGAKSLKKAISOAKAAYFPP-----NGLH--MLLGPFGS--GKSL- 156
QY 120 IILVPRHLOGRKQVGGDFLBARMSFALTAAGSKWDFNNQYIVL-----SPT 164
DB 157 ---FANRIYQFAIY-SDILKA--GAFITIFNCA---LYNNNGQLLTS-LRQHKH-SFT 205
QY 170 LFWGQVSL-----SLL---H--PSEGASALWRANOG-YIKIFKQFVN-175HV 213
DB 206 GAADKAKGLVEAGANGILFMDEIHRLEPGEQEMLEFYIDSGSYNHL---GESHKRTSNV 262
QY 217 FTEGGLTNSAELCEYLDRODAFQCMQPO--HMP-----CEALTYMTTHRE 264
DB 263 LFICATTENSSALL-----KTFIRARPMTIHIPLEERSINERVDITFLQKIAE 314
QY 265 -----VSYLIDKENSLEFRSKVG-VEMAKDR-----KIIDVTKNKKFKILETCOVAKMT 314
DB 315 RIKKNLSVHIDVYVNALHISAKFVNGVGLKSNVULCAHGFANLIDREVVELTVN----- 369
QY 314 VPGGYTLQGMKITTFCNQ-----VQDITKINGCLAKGLIYLGG 364
DB 370 DLDP--EIKQEMSSSKNMORSKAISEVYVITTIISPIYRETTIKIFHRSNRYVHLEH 427
QY 355 STLQWATYFPVPGGTYLQGMKITTENNOYDITKINGCLAKGLIYLGGSTLQWATYFP 411
DB 428 K-----VTLW-----KEGSKKDINQYIILDLVLAHSGF--HQAQVKNL 448
QY 411 YSLDHDIYPREIDRSGDKNTAIVITFGO--HFRPPIDIPIRRAIGVOKATERLEFR 467
DB 469 LTFVEDDVI-QMTQKLEIAHEIDCTFFRKIFVLSMHIDAFLEKQ---KQIVLNTQ 514
QY 468 SPARKVITKTEHREHNIETER---FGD-FNGY-----IHYLM-----KQIFK 508
DB 524 -----ETDELROTWHKEVRVAMIFKQIGYKVAIPEIYVITLMLHS-KSLKN 540

```

```

QY 509 LNWGIIADMTIAYCTDT 527
DB 576 KRQGIIV-----AHGNST 589

RESULT 3
YQ2L_YEAST
ID YQ2L_YEAST STANDARD: PRT: 1237 AA.
AC P51254:
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DE 10-MAY-2000 (Rel. 39, Last annotation update)
DE HYPOTHETICAL 140.5 KDA PROTEIN IN CTT1-PRP31 INTERGENIC REGION.
GN YGR090W.
OS Saccharomyces cerevisiae (Baker's Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_Taxid=4932;
RN 11)
RP SEQUENCE FROM N.A.
RA Medler H., Scharfe M., Medler E., Mambuli R.;
RU Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
RN 12)
RP SEQUENCE FROM N.A.
RA Hernandez K., Weber N., Wipfli P., Schmidheini T.;
RU Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
CC 1- SIMILARITY: TO S. POMBE SPCC776.08C.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CV EMBL: 272875; CAA97093.1;
DR SGD: S0003322; YGR090W.
DR Hypothetical protein.
SU SEQUENCE 1237 AA; 140484 MW; 9A2B5C8B5493D7D3 CRC64;

Query Match 3.6%; Score 103.5; DB 1: Length 1237;
Best Local Similarity 19.3%; Pred. No. 2.8; Mismatches 165; Gaps 28;
Matches 102; Conservative 87; Indels 165; Gaps 28;

QY 30 NFKLMSALNSTSVHYMNSAKSLFPTSLPLKPLTETELKEIEKLDOQIPREF 89
DB 484 NEYKTRFSINLLIGFYKVEERKILPNRNCI-----RIHQ--ESKEQSLPATPL 331
QY 40 TVAVTTTATSTATILNPRDTYCGDOLDLLEVRDHGQKQYGGDFLRAR----- 143
DB 332 YNEVSLSSTH-----ENLTK-----YLYKTKRKQTESFEVATYLGRLM 370
QY 144 -----SSPALTACAGCKVMDFNNGTYLVSFTLFWF-----GOVSLLLHPSEGASALM 193
DB 371 QVQGFSSNMHSISLQ-----GQTF--EFTLMAALLNGGINSNKLLM----- 414
QY 174 RARNOGYKTI-IFNG--KPYNGTSNHFTECGITLNSAAE-----LCGYLDDROGAFYC 244
DB 415 -----GFSSVQLEFGVILKY-ATMDLCHDGHLOQHSHPENSSSSPAASYIDEGVPTTLE 468
QY 245 MKYQHPCEALTYMTFNREV--SYLTDKENSJLFHRSKVGVEEMDR-KHIDVYNCNKE 301
DB 459 DKSTKV--NLTKMTVSSYOILKEVAGETLRLNN-----VYDQPSNIFLNISRED 519
QY 302 KLEHFGVGMKRPVPGGYTLQCKMTTFGNOVOYDITKINGCKKGLLYLGDTLRQMT 361
DB 520 NLK-----YDL-----CYDVLPLQCKYNN-LETSLATATGSMERYKEI 556
QY 362 YFFPKVATLKFDHETGIEKKHLLDAERHQIQM--KKHSYFVTEQLYS-----IL 414
DB 557 -----TLENFLAKKITVAVAYALGDRIKYIOIEWGQCKSDPITRKRYSYNTGNGHF 608

```

```

QY 415 DHDYI-----PREIDRL--SGDKNTAIVTGGHFRFPIDIFIRRAQVOKAIE-- 463
DB 609 NEDFVRKVLIVNPECECKLVTKGPAHSETMSTAFAVKNF-----MDIKSLRRPK 659
QY 464 -----LFLRSPATKVIKTEINI-----RPMHLETERGDPHGVI 497
DB 660 DGSITHCWMTSSSEPIITSSIVNFALOKHKVSKAQSINETIKKHNF 708

RESULT 4
COX1_TETPY
ID COX1_TETPY STANDARD: PRT: 698 AA.
AC P11947;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DE 01-OCT-2000 (Rel. 40, Last annotation update)
DE CYTOCHROME C OXIDASE POLYPEPTIDE I (EC 1.9.3.1).
GN COI.
OS Tetrahymena pyriformis.
OC Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Hymenostomatida;
OC Tetrahymenida; Tetrahymena.
OX NCBI_Taxid=5908;
RN 11)
RP SEQUENCE FROM N.A.
RA STRAIN-ST;
RX MEDLINE-88184706; PubMed-2833363;
RA Zlate Z., Suyama Y.;
RT "The cytochrome oxidase subunit I gene of Tetrahymena: a 57 amino
RT acid NH2-terminal extension and a 108 amino acid insert."
RL Curr. Genet. 12:357-368(1987).
CC 1- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
CC CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-
CC 3 FORM THE FUNCTIONAL CORE OF THE ENZYME. ELECTRON'S ORIGINATING IN
CC CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2
CC AND HEME A OF SUBUNIT 1 TO THE BINMETALLIC CENTER FORMED BY HEME A3
CC AND COPPER B.
CC 1- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) -> 2 H(2)O +
CC 4 FERROCYTOCHROME C.
CC 1- PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN.
CC 1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL
CC INNER MEMBRANE. CONTAINS 12 POTENTIAL TRANSMEMBRANE DOMAINS.
CC 1- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CV EMBL: L28677; AAA32102.2;
DR EMBL: X06133; NOT_ANNOTATED_CDS.
DR PIR: S00742; S00742.
DR HSP: P00396; 10CC.
DR InterPro: IPR000883;
DR Pfam: PF00115; COX1; 2.
DR PROSITE: PS00077; COX1; 1.
KW Oxidoreductase; Heme; Copper; Mitochondrion; Transmembrane;
KW Respiratory chain; Inner membrane.
FT METAL 111 111 IRON (HEME A) (PROBABLE).
FT METAL 401 401 COPPER B (PROBABLE).
FT METAL 405 405 COPPER B (PROBABLE).
FT METAL 450 450 COPPER B (PROBABLE).
FT METAL 451 451 COPPER B (PROBABLE).
FT METAL 536 536 IRON (HEME A3) (PROBABLE).
FT METAL 538 538 IRON (HEME A) (PROBABLE).
SU SEQUENCE 698 AA; 81746 MW; D6B0F48BA3109A72 CRC64;

```

Query Match 3.4%; Score 100; DB 1; Length 698;
Best Local Similarity 19.2%; Pred. No. 2.4;
Matches 105; Conservative 81; Mismatches 194; Indels 166; Gaps 30;

```

OY 23 MFIISQNTKLM-----SALNLSISVH---YMNNSKSLFPRKSLIPKLETELRLK 74
DB 1 MLYIFN-SFDMWYDFLEOTKSPKSVNNYITNLIKLF---TYLNDLR-----K 48
OY 75 EIEKLDQOIPPRF-----THVNTTSAITSTAITLPDHYCAGDQDILLEY 124
DB 49 HILKRYTTINHKRIAINLYFSWVTGLSGAALATWIRLELHNPSPFPFGDLSRLQVY 108
OY 125 RDHGORROYGDFLRARMSPALTAGASGVMDPNNGYLSFTL-----FNE---GCV 176
DB 109 TAH-----GLIMVFYVYPILEGGFANFLPIYVSGKDAVPRNLSIGFIOPGYI 160
OY 177 SLSTLL-LIHPSGASALMARNOGYDKIIFKGFVNGTSHVTECGTLNSNALECEYLD 235
DB 161 LIAKIGFLRPQ-----FWRY---YDKTSFSPFLEKMY-----NQYKEKN 199
OY 236 D-----RDGEAFY-----CMKPOH-----MPCALTYMTRNRKREVSILTKE 272
DB 200 DYLEFYLDELKKEITDHPFKMARVYKLPQYVSFPLKLMKMTWIPESF-----254
OY 273 NSLFRSKVGEVEMKDKRKIDVYTCNKRKIEETCOVGMKPPVGGYTLQGMITTFGNO 332
DB 255 --WYASRV---VOSRRKKVEFTKCSARLT-----TTAGWT---FITPFS- 292
OY 333 VOLDTIKINGCLKGKLIYLLDSTLRQMIYFPKYVTKLFKFDLHEGIFKKHLLDAER 392
DB 293 ---NIKYTG-----VGSQDILISVFAGSTISTITNL---ITRNTLMPQMR 336
OY 393 HTOIQMKKSHYPVTFOLYSLIDHDYIPREIDRLSGDKNNAIV-ITFGGHERPFIIDFI 451
DB 337 HRVY-----LMFVYISIFELT-----MLMTITPVGAIVTMAFDRHWQ---TTFF 381
OY 452 RRAIGVQKAIER---LFLSPATKVIITKE---NIREMHETERGDPH-----G 495
DB 382 EYAGGDPIILSOHLFWFGHREYVLLIPTFGFINMIVPNNRRARASHHMAIYVMA 441
OY 496 YIHLYI 501
DB 442 YMGYLV 447

```

RESULT 5
YJ37_CAEL
ID YJ37_CAEL STANDARD; PRT; 400 AA.
AC P34428;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-JUN-1994 (Rel. 29, Last annotation update)
DE HYPOHETICAL 45.5 KDA PROTEIN F44B9.7 IN CHROMOSOME III.
OS F44B9.7
OS Caenorhabditis elegans.
CC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditiida; Rhabditiidae;
CC Rhabditiidae; Pelodermidae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Almscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favell A., Fraser A.,
RA Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,
RA Johnston L., Jones A., Kerhaw J., Kirsten J., Lalister L., Jier M.,
RA Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifkin L., Roopia A., Saunders D., Showkeen R.,
RA Sims M., Smalton N., Smith A., Smith M., Sonhammer E., Staden R.,
RA Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,
RA Watson R., Watson A., Weinstock L., Wilkinson-Sproat J.,
RA Woldman P.;

RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sdb.ch/about/try/>
CC or send an email to license@isb-sdb.ch).

DR EMBL: I23648; AAA28029.1; -
DR PIR: S44810; S44810.
DR Wormpep: F44B9.7; CE00176.
KW Hypothetical protein.
SQ SEQUENCE 400 AA; 45523 MW; 96BF6BC3B4C2D7D CMC64;

Query Match 3.4%; Score 99.5; DB 1; Length 400;
Best Local Similarity 22.0%; Pred. No. 1.2;
Matches 70; Conservative 36; Mismatches 129; Indels 81; Gaps 1;

```

OY 56 PKISLT-----PLKPLETELRIKELIEKLDQOIPRPETHVNTTSAITSAI 105
DB 62 PTTSLAQHLSSAPLPAPPOOOSHL-----SGLDQPTTSTAINHAPLOTI 111
OY 106 LNPEDTYCGDQDILLEYRDHUGORXOYCGDFLRARSSALTAASAKVDFPNNGYL 105
DB 112 ARREDNEMANQDLBELAARQ--QOQET--MLDQHQOQONRBAQDLAEVAFAN---104
OY 166 VSTLEWEOVSLILLIHPSGASALMARNOGYDKIIFKGFVNGTSHV---FTFCG 221
DB 163 -----PPTYSIPPOGTGHPPOOTSASSOURAOE-----TMAQOTPGUSHIHESQNETN 211
OY 222 LTNSNAELC-----EYLDPRDGEAFYC-----MKPOHM-----PTEALTYM 274
DB 211 LEKYSTGELCYGRELVDLVNTVSTLSITLKVMKRPINQENENDLAQKNAQMR 270
OY 259 TTRNREVSYLDKNSLPHSKYG---VENMKD---RKHIDVYTCNKRKIEE--TGVV 409
DB 271 S-----DIQIIEKREPTWKRMTGEDYIELMLDSELKPKMDEKRRKRLLOJFAIVI 326
OY 310 GMRPVPGSYTLQCKWIT 327
DB 327 SKTPPEGHIMYQCKWIS 344

```

RESULT 6
SPAT_BACSU
ID SPAT_BACSU STANDARD; PRT; 614 AA.
AC P93116;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE SUBTILIN TRANSPORT ATP-BINDING PROTEIN SPAT.
GN SPAT OR SPAB OR SPAY.
OS Bacillus subtilis.
CC Bacteria; Firmicutes; Bacillus/Clostridium group;
CC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 6633;
RX MEDLINE=92171481; PubMed=1539969;
RA Klein C., Kalletta C., Schnell N., Entian K.-D.;
RT "Analysis of genes involved in biosynthesis of the antibiotic
RT subtilin.";
RL Appl. Environ. Microbiol. 58:132-142(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 6633;
RX MEDLINE=92138640; PubMed=1735728;

```

RA  Chung Y.J., Steen M.T., Hansen J.N.;
RT  "The subtilin gene of Bacillus subtilis ATCC 6633 is encoded in an
RT  operon that contains a homolog of the hemolysin B transport
RT  protein."
RL  J. Bacteriol. 174:1417-1422(1992).
CC  -1- FUNCTION: PROBABLY IMPLICATED IN THE EXPORT PROCESS OF THE
CC  LANTIBIOTIC SUBTILIN.
CC  -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN
CC  -1- SIMILARITY: BELONGS TO THE ATP-BINDING TRANSPORT PROTEIN FAMILY
CC  (ABC TRANSPORTERS). HLYB SUBFAMILY.
CC  -----
CC  THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL: M86869; AAA22838.1; -.
DR  EMBL: M83944; AAA22770.1; ALT_INIT.
DR  EMBL: M9263; AAA22776.1; ALT_INIT.
DR  EMBL: U09819; AAB91587.1; -.
DR  PIR: B43935; B43935.
DR  HSSP: P13569; 1NRD.
DR  InterPro: IPR001617; -.
DR  Pfam: PF00005; ABC_tran; 1.
DR  PROSITE: PS00211; ABC_TRANSPORTER; 1.
KW  Transport; protein transport; Bacteriocin transport; Transmembrane;
KW  ATP-binding.
FT  NP_BIND 387 394  ATP (POTENTIAL).
FT  CONFLICT 39 39  R -> A (IN REF. 2).
FT  CONFLICT 68 68  E -> D (IN REF. 2).
FT  CONFLICT 394 394  K -> S (IN REF. 2).
FT  CONFLICT 407 408  HE -> QC (IN REF. 2).
SQ  SEQUENCE 614 AA: 71188 MW: 84262EAD036C33F CRC64;

```

Query Match 3.44; Score 99.5; DB 1; Length 614;
 Best Local Similarity 20.44; Pred. No. 2.2;
 Matches 128; Conservative 88; Mismatches 212; Indels 199; Gaps 29;

```

QY  5 TMLQKTLILISFSVYTWMTIIS-----QNTK 33
   : : : : : : : : : : : : : : : : : :
DB  28 TLEKSLFLKIRFSITGILPVSLSQELINSVLRKEVSIVITFLYLGVSFSE 87
   : : : : : : : : : : : : : : : : : :
QY  34 LWS-----ALNLSIVHWNNNSAKSLPQKSLPLKPLTELEIKELIKLDDQ 83
   : : : : : : : : : : : : : : : : : :
DB  88 LLSQISEFYKCKPQNLIGYKLY-----KYMKSSNLALDFENPEIYDK-LERYTKE 139
   : : : : : : : : : : : : : : : : : :
QY  84 IPRPFTHVTTTSATHTATIL-----NPRDTYCRDDQDILLERYD----- 126
   : : : : : : : : : : : : : : : : : :
DB  140 ISYKPOIIOAITMTSFTVLLSLAFLMSWPK-----VSLLLVLPVLSLYFL 191
   : : : : : : : : : : : : : : : : : :
QY  127 HLQGRQY-----GGDFLRAMSSPALTAGASQYWD-FNNGTYLVSFTLFWGGVSL- 178
   : : : : : : : : : : : : : : : : : :
DB  192 KIGGEFFHMKKCKEKESWYSYILTHDFSPKELKYNLKDLYLN-KYWDIKSFIE 249
   : : : : : : : : : : : : : : : : : :
QY  179 -----SLLIHPS-----GASLMRARNQGDILIKKGFY-----NGTS 214
   : : : : : : : : : : : : : : : : : :
DB  250 QUTKILRKKTLLNLYETAVOLGAVITFLPAISAGAKINGWNSYINSVLYONSSQ 309
   : : : : : : : : : : : : : : : : : :
QY  215 HVTGEGILINSNA---ELCEYLDLRDQEAFCYKQHPHMEALTYTTRNREVSYLDK 271
   : : : : : : : : : : : : : : : : : :
DB  410 SIMTSIYSINSLYNNQJYEFLELKEKESQGHKPIVEPIHSYVF---QNSFYYPNG 366
   : : : : : : : : : : : : : : : : : :
QY  272 FNSLFRHSKYGVEMKDRKHIDVYCNCKRKEIETQYGMKPPVPGYTLQGR---RT 327
   : : : : : : : : : : : : : : : : : :
DB  467 EGLT-----KHINVS-----LAKGRVAIVG-----PNG---SGKKTIFKL 400
   : : : : : : : : : : : : : : : : : :
QY  424 FPCNOVOLGTLKNGCLKGKLYILIGSTLRQWLYFFPKVYK---TLK-----PFDLH 377
   : : : : : : : : : : : : : : : : : :
DB  401 TGLYEVHGGDILNIGI---NIKELMDMSYNNQJALFQDFKMYEMTLKENIGFGQIDKIH 457

```

```

QY  378 ETGIFKKHLLDAER-----ETQI-----QMKKHSYPTVTPOLYS 412
   : : : : : : : : : : : : : : : : : :
DB  458 QTN--KMEHEVDIYRADELKSHSSYQDFDGLWFDGROLSGGQWQKIALARAYFREAS 515
   : : : : : : : : : : : : : : : : : :
QY  413 LIHDHYPEIDRLSGDKNFAIVITFGHF--RPFIDIFIRFAIVQKATERFLFSPSA 470
   : : : : : : : : : : : : : : : : : :
DB  516 LYIDPSSALDPIAENE-----TDFTFSLSKDKIGITISHRLVAAKLADRIIVMDK 569
   : : : : : : : : : : : : : : : : : :
QY  471 TKV-----IKTENIREKHIEERF 490
   : : : : : : : : : : : : : : : : : :
DB  570 EIVGIGHHELLKTCPIYKKWDESNV 596
   : : : : : : : : : : : : : : : : : :

```

RESULT 7
 SCAL_RAT STANDARD: PRT: 698 AA.

```

AC  P37089; 064593;
DT  01-OCT-1994 (Rel. 30, Created)
DT  01-OCT-2000 (Rel. 40, Last sequence update)
DT  01-OCT-2000 (Rel. 40, Last annotation update)
DE  AMILORIDE-SENSITIVE SODIUM CHANNEL ALPHA-SUBUNIT (EPITHELIAL NA+
DE  CHANNEL, ALPHA SUBUNIT) (ALPHA ENAC) (NONVOLTAGE-GATED SODIUM CHAN-
DE  EL 1 ALPHA SUBUNIT) (SCNEN) (ALPHA NACH).
GN  SCN1A OR RENAC.
OS  Rattus norvegicus (rat).
OC  Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC  Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX  NCBI_Taxid=10116;
   [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=Wistar; TISSUE=Distal colon;
RX  MEDLINE=93170495; PubMed=8381272;
RA  Lingueglia E., Volley N., Waldmann R., Lazdunski M., Barbry P.;
RT  "Expression cloning of an epithelial amiloride-sensitive Na+ channel.
RT  A new channel type with homologies to Caenorhabditis elegans
RT  degenerins."
RL  FEBS Lett. 318:95-99(1993).
   [2]
RP  SEQUENCE FROM N.A.
RC  STRAIN=SPRAGUE-DAWLEY; TISSUE=Colon epithelium;
RX  MEDLINE=93156815; PubMed=8381523;
RA  Canessa C.M., Horisberger J.D., Rossier B.C.;
RT  "Epithelial sodium channel related to proteins involved in
RT  neurodegeneration."
RL  Nature 361:467-470(1993).
   [3]
RP  SEQUENCE FROM N.A.
RC  STRAIN=Wistar-Kyoto; AND SHRSF; TISSUE=Kidney;
RX  MEDLINE=97191134; PubMed=9039092;
RA  Kreutz R., Strub B., Rubattu S., Hubner N., Szpirer J., Szpirer C.,
RT  Ganem D., Lindpaintner K.;
RT  "Role of the alpha-, beta-, and gamma-subunits of epithelial sodium
RT  channel in a model of polygenic hypertension."
RL  Hypertension 29:131-136(1997).
   [4]
RP  TOPOLOGY, AND CARBOHYDRATE-LINKAGE SITES.
RX  MEDLINE=95014183; PubMed=7929098;
RA  Snyder P.M., McDonald F.J., Stokes J.B., Welsh M.J.;
RT  "Membrane topology of the amiloride-sensitive epithelial sodium
RT  channel."
RL  J. Biol. Chem. 269:24379-24383(1994).
   [5]
RP  TOPOLOGY.
RX  MEDLINE=94230383; PubMed=8175716;
RA  Renard S., Lingueglia E., Volley N., Lazdunski M., Barbry P.;
RT  "Biochemical analysis of the membrane topology of the amiloride-
RT  sensitive Na+ channel."
RL  J. Biol. Chem. 269:12981-12986(1994).
   [6]
RP  MUTAGENESIS OF SER-569 AND SER-593.
RX  MEDLINE=95263507; PubMed=7744818;
RA  Waldmann R., Champigny G., Lazdunski M.;

```

```

RT      *Functional degenerin-containing chimeras identify residues essential
RT      for amiloride-sensitive Na+ channel function.*;
RT      J. Biol. Chem. 270:11735-11737(1995).
CC      -1- FUNCTION: SODIUM PERMEABLE NON-VOLTAGE-SENSITIVE ION CHANNEL
CC      INHIBITED BY THE DIRECTIC AMILORIDE. MEDIATES THE ELECTRODIFFUSION
CC      OF THE LUMINAL SODIUM (AND WATER, WHICH FOLLOWS OSMOTICALLY)
CC      THROUGH THE APICAL MEMBRANE OF EPITHELIAL CELLS. CONTROLS THE
CC      REABSORPTION OF SODIUM IN KIDNEY, COLON, LUNG AND SWEAT GLANDS.
CC      ALSO PLAYS A ROLE IN TASTE PERCEPTION.
CC      -1- SUBUNIT: HETEROTETRAMER OF TWO ALPHA, ONE BETA AND ONE GAMMA
CC      SUBUNIT. A DELTA SUBUNIT CAN REPLACE THE ALPHA SUBUNIT.
CC      -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC      -1- SIMILARITY: BELONGS TO THE AMILORIDE-SENSITIVE SODIUM CHANNELS
CC      FAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation.
CC      The European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL: X70521; CAA49916.1; ALT_INIT.
DR      EMBL: X70497; CAA49905.1; -.
DR      EMBL: U54699; AAB61156.1; -.
DR      EMBL: U54700; AAB61157.1; -.
DR      PIR: S29499; S29499.
DR      InterPro: IPR001873; -.
DR      Pfam: PF00858; ASC: 1.
DR      PRINTS: PR01076; AMINACHANNEL.
DR      PROSITE: PS01206; ASC: 1.
KM      Ionic channel; Transmembrane; Ion transport; Glycoprotein.
FT      DOMAIN 1 110
FT      TRANSMEM 1 131
FT      DOMAIN 132 589
FT      TRANSMEM 590 610
FT      DOMAIN 611 698
FT      CARBOHYD 190 190
FT      CARBOHYD 259 259
FT      CARBOHYD 320 320
FT      CARBOHYD 339 339
FT      CARBOHYD 424 424
FT      CARBOHYD 538 538
FT      MUTAGEN 588 588
FT      MUTAGEN 592 592
FT      CONFLICT 598 599
FT      SEQUENCE 698 AA; 78887 MW; B0CF7C1563CE9763 CRC64;
Query Match 3.44; Score 98; DB 1; Length 698;
Best Local Similarity 19.28; Pred. No. 3.5; Mismatches 157; Indels 164; Gaps 18;
Matches 89; Conservative 53;
OY      11 LLIISFVYVWIPISQNTFKMLGALNLSIVHWNNKSAKLPKTSILPLKPLTELE 70
OY      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      114 VLMLCTFGMMYQFALLEEYLYSPVSLNIN-----NSDKLVPAVYVCLNPRYTE 167
DB      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY      71 LRIKEIKLDO-----QIPRP 89
OY      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      168 --IKEELFELDRIITBOTLFDLYKNSSTYRQAGARRSSDGLGAPPHLORLKRPP 225
DB      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY      90 THWTTTSATHTATILNPR-----DTYGRDQDILLLEVRDH 127
OY      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      226 S--GRTARSGSSSVNPNPDRDKKICFOLCQNNKSDCFYOTSSG--VDAYRE---- 277
DB      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY      128 LQGRKQYGGDFLRAMSSPALTAGASGKMDPNNGTIVSEFTLFWEGGYSLILLHPSE 187
OY      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      278 -WYRFHYINILSRISDTSPLALEEALG-----NFTICFPNAPRQANANIS-KFHHPK 329
DB      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

OY      168 G-----ASLRPANOQYDKILFKGPNFNGSHWTECG::LNSNLSCEYLDDRD 248
OY      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      330 GNCYFENDKNNKNSMLSSMPG-----VNN-----SLSLIRIEQNFY---- 447
DB      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY      239 QEAFCYCKPOHMFCEALTYMTTRNRESVLTDKESLPHKNSGVEMMKIKHHII VTN 408
OY      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      368 -----PLSTVIGARVYVHQDEPARMDQGNLHGVENSISKKEA--LUSLQGN 417
DB      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY      299 KREKIEETQYGMKRPVPGVGYTTLQGMKITTFCNOVOLDI KINSGCKKILYLLDSTL 346
OY      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      418 YGDTENGSDVPVKNLPSKYTDQ-----VCHSCFENMKIKKCA--- 454
DB      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY      359 QWITY-FPKVYKTLKF-----FDLHETGIFKK 384
OY      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      460 -YIFPYPKGVFCDFYKROSSWGYCYKLGAFSLDSIGCFSK 501
DB      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
RESULT 8
HUI4_YEAST STANDARD; PRT; 892 AA.
ID HUI4_YEAST
AC P40985;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE PROBABLE UBIDUITIN--PROTEIN LIASE HOLD 4 (Ec 6.3.2.-).
GN HUI4 OR YNR036C OR J1608.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C;
RX MEDLINE=95397595; PubMed=7668047;
RA Huang M.-E., Chait J.-C., Galibert F.;
RT "Analysis of a 42.5 kb DNA sequence of chromosome X reveals three
RT tRNA genes and a new open reading frames including a gene most
RT probably belonging to the family of ubiquitin-protein ligases."
RL Yeast 11:775-781(1995).
RP [2]
RP SEQUENCE OF 362-892 FROM N.A.
RX MEDLINE=95045431; PubMed=7957102;
RA van Gool A.J., Verhage R., Swaenmakers S.M.A., van de Putte P.,
RA Brouwer J., Troelstra C., Bootsma D., Hoelijmakers J.H.J.;
RT "Rad26, the functional S. cerevisiae homolog of the Cockayne syndrome
RT B gene ERCC6."
RL EMBO J. 13:5361-5369(1994).
RP [3]
RP GENE NAME, AND GENE DISRUPTION.
RX MEDLINE=9907972; PubMed=9858558;
RA Wang G., Yang J., Hultberg J.M.;
RT "Functional domains of the raps ubiquitin-protein ligase."
RL Mol. Cell. Biol. 19:342-352(1999).
OY      -1- FUNCTION: PROBABLE E3 UBIDUITIN-PROTEIN LIASE WHICH ACTS
OY      UBIDUITIN FROM AN E2 UBIDUITIN-CONUGATING ENZYME IN THE FORM OF A
OY      THIOESTER AND THEN DIRECTLY TRANSFERS THE UBIDUITIN TO TARGETED
OY      SUBSTRATES. NON ESSENTIAL.
OY      -1- MISCELLANEOUS: A CYSTEINE RESIDUE IS REQUIRED FOR
OY      UBIDUITIN-THIOLESTER FORMATION (BY SIMILARITY).
OY      -1- SIMILARITY: CONTAINS 1 HECT-TYPE E3 UBIDUITIN-PROTEIN LIASE
OY      DOMAIN.
OY      -----
OY      This SWISS-PROT entry is copyright. It is produced through a collaboration
OY      between the Swiss Institute of Bioinformatics and the EMBL outstation.
OY      The European Bioinformatics Institute. There are no restrictions on its
OY      use by non-profit institutions as long as its content is in no way
OY      modified and this statement is not removed. Usage by and for commercial
OY      entities requires a license agreement (see http://www.isb-sib.ch/announce/
OY      or send an email to license@isb-sib.ch).
OY      -----
DR      EMBL: L76344; AAA88738.1; -.
DR      EMBL: Z49516; CAA89563.1; -.

```

```

DR EMBL: X81635; CAAS7291.1;
DR PIR: S48296; S48296.
DR SGD: S0003797; XJ036C.
DR InterPro: IPR000569; -.
DR Pfam: PF00632; HECT. 1.
DR PROSITE: PS50237; HECT. 1.
KM Dblquitin conjugation: Ligase.
FI DOMAIN 792 892
FT BINDING 860 860
FT CONFLICT 362 362
FT CONFLICT 410 410
FT CONFLICT 418 418
FT CONFLICT 514 519
FT CONFLICT 589 607
FI R (IN REF. 2)
FI CONFLICT 705 705
FI CONFLICT 723 724
FI SEQUENCE 892 AA; 103456 MW; 35E9VARDVA92HD84 CRC64;

Query Match
Best Local Similarity 3.4%; Score 98; DB 1; Length 892;
Matches 53; Conservative 53; Mismatches 85; Indels 90; Gaps 12;

UY 114 RGQDLDLLEVR-----IHGQKKYQGDLELR-----MSSPALITGASGYMDFNN 161
DB 513 KQKSVAVYFKIKVRDVISHSLSKCIKEHQGLKSLRIEFVNEPQIDAGLRKEMF-- 570
UY 162 GIVAVSTFLMEGQVSLSLIHPSSEASALWRAHNGYDKILFKGK----- 208
DB 571 ---LTKSLF-----NPMNGLFITYIKSSRSWPAIDPPAFQK--SKGNSQLDLYLFGV 621
UY 209 ----FVNGT-----SHVFEGCLTINSALCEYDDROE 240
DB 622 MGLATFNSTLDLQPKALYKLCSEPLSEFSEPE--TSRLTKMLWTFEDNED 678
UY 241 AFYCKMPOHPCPEALTYMTNRNREVSYLTDKENSLSFHSKVGVEKMKRKHIDVTCNKR 300
DB 679 VF-----SLTFETTYRNMMILNDSKRS---KEYATVELCNGSNVPIQTSKHK 724
UY 701 EKIHETQYCKMKPPVPGYTIQGWITTF-----CNGVOL 335
DB 725 EFVKKWVEFLERKSTLEPOL--NKEVSGRKRVPAECNSIKL 762

RESULT 9
ID THDE_RAT STANDARD: PRT: 1025 AA.
AC 01-OCT-2000 (Rel. 40, Created)
UT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE THYROTROPIN-RELEASING HORMONE DEGRADING ECOTOENZYME (EC 3.4.19.6) (TRH-
DE DEGRADING ECOTOENZYME) (TRH-DE) (TRH-SPECIFIC AMINOPEPTIDASE)
DE (THYROTROPIN-RELEASING HORMONE DEGRADING ECOTOENZYME) (TRH-DE) (TRH-
GN TRHDE.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=PIUITARY;
RX MEDLINE=95023946; PubMed=937801;
RA Schander B., Schenburger L., Koehle J., Bauer K.;
RT "Cloning of a cDNA encoding an ectoenzyme that degrades thyrotropin-
RT releasing hormone.";
KL Proc. Natl. Acad. Sci. U.S.A. 91:9534-9538(1994).
CC -!- FUNCTION: SPECIFIC INACTIVATION OF TRH AFTER ITS RELEASE
CC -!- CATALYTIC ACTIVITY: RELEASE OF THE N-TERMINAL PRODOTAMYL GROUP
CC FROM PGDL-1-HIS-XAA TRIPEPTIDES AND PGDL-1-HIS-XAA-GLY
CC TETRAPEPTIDES.
CC -!- COFACTOR: BINDS ONE ZINC ION (BY SIMILARITY).

```

```

CC -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN.
CC -!- TISSUE SPECIFICITY: PREDOMINANTLY EXPRESSED IN BRAIN AND
CC PITUITARY. LOWER LEVELS IN LUNG AND LIVER.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M1 (ZINC METALLOPROTEASE);
CC ALSO KNOWN AS THE PEPT SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sdb.ch/announce/
CC or send an email to license@sdb.ch).
CC -----
DR EMBL: X80535; CAAS6675.1; -.
DR MEMOPS: M01.008; -.
DR InterPro: IPR000130; -.
DR InterPro: IPR001930; -.
DR Pfam: PF01433; Peptidase_M1; 1.
DR PRINTS: PR00756; ALADIPASE.
DR PROSITE: PS00142; ZINC_PROTEASE; 1.
DR HydroLase: Metalloprotease; Aminopeptidase; Zinc; Glycoprotein;
KM Transmembrane; Signal-anchor; Sulfatation; Phosphorylation.
FT DOMAIN 1 40
FT TRANSMEM 41 61
FT SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT (POTENTIAL).
FT METAL 62 1025
FT METAL 441 441
FT ACT_SITE 442 442
FT METAL 445 445
FT METAL 464 464
FT ACT_SITE 528 528
FT MOD_RES 30 30
FT MOD_RES 381 381
FT CARBOHYD 90 90
FT CARBOHYD 161 161
FT CARBOHYD 176 176
FT CARBOHYD 223 223
FT CARBOHYD 339 339
FT CARBOHYD 606 606
FT CARBOHYD 635 635
FT CARBOHYD 650 650
FT CARBOHYD 664 664
FT CARBOHYD 685 685
FT CARBOHYD 801 801
FT CARBOHYD 907 907
SQ SEQUENCE 1025 AA; 117286 MW; 4024EB26208D16B CRC64;

Query Match
Best Local Similarity 3.4%; Score 98; DB 1; Length 1025;
Matches 64; Conservative 48; Mismatches 96; Indels 92; Gaps 16;

UY 151 GASCKVMDFNNGIYVSTFLMEGQVSLSLIHPSSEAS--ALMRARNGYDKILF--K 206
DB 625 GAKTKALQLONSSYL-----W--QIPLTIVGNKSHVSSAIIWVSNKSEHRTIYDK 676
UY 207 GKFNQTSHYETECG-----LTLSNMLCEYLDHDDQDAF 242
DB 677 GSWLGN--INQGVFRVNYDLNRWLLDQLIRNHEVLSVSRAL--IDDAFSLAR 730
UY 243 YCKMPOHPCPEALTYMTNRN-----REVSLTDC-----EN-SLFR----- 278
DB 731 AGYLPQNIPELIRYVSEKQPLFMHAAASALYPLDKLLDMENYINFEYILKQVATY 790
UY 279 SKVGEVEMKDKRHIDVTNQN-----KREIEETGVCKMPPVPGYTIQGWITTFPCNQ 332
DB 791 SKLG--NPKNNFNGSVQASQHELELRREYIMLASPFKNCHQOASTLSISSNNR 848
UY 333 VQDLTININCCLKKGIYLDGSLTRQ--MIYFPKVKYTKLFPDLHETGIFKKLLDA 390
DB 849 IPLNVDI-----VICTGSLDDEWEMFMKFNST-----TAVSEKILLDA 892

```


90	SEQUENCE		1121 AA; 129668 MW; E507095f296AED03 CRC64;
89	KM	DNA REPAIR.	
88	RW	Hydrolase; Nuclease; Exonuclease; Endonuclease; Helicase;	
87	DR	TIGR: H10942;	
86	CC	EMBL: U32775; AAC22596.1; -	
85	CC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration	
84	CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation.	
83	CC	The European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way	
82	CC	modified and this statement is not removed. Usage by and for commercial	
81	CC	entities requires a license agreement (See http://www.isb-sdb.ch/announce/	
80	CC	or send an email to license@isb-sdb.ch).	
79	CC	- SUBUNIT: CONSIST OF THREE SUBUNITS; RECB, RECC AND RECD	
78	CC	(BY SIMILARITY).	
77	CC		
76	CC		
75	CC		
74	CC		
73	CC		
72	CC		
71	CC		
70	CC		
69	CC		
68	CC		
67	CC		
66	CC		
65	CC		
64	CC		
63	CC		
62	CC		
61	CC		
60	CC		
59	CC		
58	CC		
57	CC		
56	CC		
55	CC		
54	CC		
53	CC		
52	CC		
51	CC		
50	CC		
49	CC		
48	CC		
47	CC		
46	CC		
45	CC		
44	CC		
43	CC		
42	CC		
41	CC		
40	CC		
39	CC		
38	CC		
37	CC		
36	CC		
35	CC		
34	CC		
33	CC		
32	CC		
31	CC		
30	CC		
29	CC		
28	CC		
27	CC		
26	CC		
25	CC		
24	CC		
23	CC		
22	CC		
21	CC		
20	CC		
19	CC		
18	CC		
17	CC		
16	CC		
15	CC		
14	CC		
13	CC		
12	CC		
11	CC		
10	CC		
9	CC		
8	CC		
7	CC		
6	CC		
5	CC		
4	CC		
3	CC		
2	CC		
1	CC		

Query Match	3.44	Score 98	DB 1	Length 1121
Best Local Similarity	18.93%	Ped. No. 6, 8		
Matched 102	Conservative	97	Mismatches 204	Indels 136
				Gaps 31
QY	49	NSASLPPKPTSLIPKPLTEELRIKELIEKLDQQLPPPEFLHNTTTSAT---	HSTAT 104	
Db	411	NQDPSLPPDYYVMADINQTYTYIOAFVFGKNGDVPQIPFSLDNKLSDDVLSVLT	470	
QY	105	ILNRPDYRGQQLDILLEYRDLHGKQKQY--GD--FLARHSSPALTAGASKY--M	157	
Db	471	LIRKESNLSADVLVLTLDIP--AMRRFNLISLADPLVREAWTDSGIFRFGKQNGDI	527	
QY	158	DFNNGTYLVSTLPEWQGVSLILLIHSEGCASALWR--ARNQGYKLIIFRGRVNGTS	214	
Db	528	NFNS-----WQAGLE-RMLIGYAMREEGIIWQDLSGLNYSV---LKGELAGLS	573	
QY	215	HVFTEGLTLNSMALECEYLADRDGEAYLCMKRQHPHCALITYTTTPREWS----	YLTG 270	
Db	574	HFEFLTALSA-----LHETLQQAHSLEKKQDELTRALLSLDPFRANEDYSDMLFPIQE	622	
QY	271	KENSLPFRHSKGVQEMKDKRKHIDVTNNCKR---EKLEETQYQCMKPPVDVGGITLQGWIT	327	

```

Dr 623 KINEL-----AEHLKTLFNEELQAEVADVIIMTCLLAPSLKFLAGK-V 667
Qy 328 TFCNOVOLDITINGCLKGLKLYLLIGDSITLKHQIITYFRVVKTLAKFDLHETGLK---- 348
Db 668 NPTCLPLMNSVVF-----KVCCLGMDAD-----YPR-TQIFNSLTMJYHYKLNHNV 72
Qy 384-----KHLLDAENHTQIQMKRHSYFVDTOLSLIDHYTFRE-----LRLS-CC 429
Db 716 RRDDEKLELEH-----LLAARDCYISVGRSITDNO-PKRPANVLSJLIIYIMVGO 767
Qy 430 KNTAIVITFGCH-FRPPIDITIRRAIGVQKAIERLFLKSPAK----- 472
Db 768 KEN-VLTIETPMTASPDNF-----KNNEK-FTRSPATKMLPLAUFUASSNNSF 816
Qy 473 VIKTINIEEM-HETEREGDF-HGYTHYLMKD-----FKIMVNLILAMTAYIL 525
Db 817 AVYMTENLEKIEEVELDALVSEVNPVKFFEGQGVYFHOKERLADSENVTLIS-GLD 874

RESULT 11
NPH3_HUMAN
ID NPH3_HUMAN STANDARD; PRT; 221 AA.
AC 095157; 090LR1;
DT 01-OCT-2000 (Rel. 40, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE NEUROKOPHILIN 3 (FRAGMENT)
GN NMPH3 OR NPH3 OR KIAA1159.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RA MEDLINE-20039618; PubMed-10574461;
RA Hironaga M., Nagase T., Ishikawa K.-I., Kikuno R., Nomura N.,
RA Ohara O.;
RT "Characterization of cDNA clones selected by the Genemark analysis
RT from size-fractionated cDNA libraries from human brain.";
RL DNA Res. 6:329-336(1999).
RN [2]
RP SEQUENCE OF 49-221 FROM N.A.
RC PubMed-9570794;
RA Missler M., Suedhof T.C.;
RT "Neurokophilins form a conserved family of neuroprotein-like
RT glycoproteins.";
RL J.Mol. Neurosci. 18:3630-3638(1998).
CC -1- FUNCTION: MAY BE SIGNALING MOLECULES THAT RESEMBLE NEUROPEPTIDES
CC AND THAT ACT BY BINDING TO ALPHA-NEUREXINS AND POSSIBLY OTHER
CC RECEPTORS (POTENTIAL).
CC -1- SUBCELLULAR LOCATION: SECRETED (POTENTIAL).
CC -1- TISSUE SPECIFICITY: BRAIN-SPECIFIC.
CC -1- SIMILARITY: BELONGS TO THE NEUROKOPHILIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation.
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcv/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: AB032885; BAA66473.1; -;
DR EMBL: AF043468; AAO02281.1; -;
DR MIM: 604636; -.
KM Glycoprotein.
FT NON_TER 1
FT CARBOHYD 31 1
FT CARBOHYD 96 96
FT CARBOHYD 106 106
FT CARBOHYD 112 112
FT CARBOHYD 112 112
N-LINKED (GLCNAC. . .) (POTENTIAL).
N-LINKED (GLCNAC. . .) (POTENTIAL).
N-LINKED (GLCNAC. . .) (POTENTIAL).
N-LINKED (GLCNAC. . .) (POTENTIAL).

```

SEQUENCE 221 AA; 24768 MW; 19AF0BF4C12CAB72 CRC64;

Query Match

Best Local Similarity 23.38; Pred. No. 0.87;

Matches	35;	Conservative	32;	Mismatches	33;	Indels	26;	Gaps	0;
---------	-----	--------------	-----	------------	-----	--------	-----	------	----

cy	189	-----ASALMARNOGYDKITIFKGFVNGTSVHTFCGTLTNSNLCCEYLD	235
nb	125	ETHDQGLFTEAKNAKSTFNCNFE-WEK-VERGR-----PTSLCTDPAKLSGDH	172
cy	236	DRDQAFYQKRPQHMCECALTMTTRNREV	265
nb	173	AUSATMCSGUPKRVCCVITATSYDHYLV	202

RESULT 12

NFH4_RAT		NFH3_RAT		STANDARD;		PRT;		252 AA	
ID		ID							

PT 01-OCT-2000 (Rel. 40, Created)
PT 01-OCT-2000 (Rel. 40, last sequence update)
PI 01-OCT-2000 (Rel. 40, last annotation update)
PE NEUROPHILIN 3 PRECURSOR.
GN NXPPI or NPPI
GS *Rattus norvegicus* (Rat)
CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Ratus
NCB TaxId=10116;

EN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-Brain:
 RC PubMed:9570794:
 RA Messing M., Suedhof T.C.:
 RT "Neurophilins form a conserved family of neuropeptide-like
 KL glycoproteins."?
 J. Mol. Neurosci. 18:3630-3638(1998).

CC - FUNCTION: MAY BE SIGNALING MOLECULES THAT RESEMBLE NEUROPEPTIDES
 CC AND THAT ACT BY BINDING TO ALPHA-NEUREXINS AND POSSIBLY OTHER
 CC RECEPTORS (POTENTIAL).
 CC -1- SUBCELLULAR LOCATION: SECRETED (POTENTIAL).
 CC -1- SIMILARITY: BELONGS TO THE NEUREXOPHILIN FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

EMBL: AF042713; AAD02226.1; -

Sample	Glucoprotein: Signal	POTENTIAL
KM	1	NEUREXOPHILIN 3
FT	252	
FT	?	
FT	62	N-INNED (GLCANC
FT	62	.) (POTENTIAL)
FT	127	N-INNED (GLCANC
FT	127	.) (POTENTIAL)
FT	137	N-INNED (GLCANC
FT	137	.) (POTENTIAL)
FT	143	N-INNED (GLCANC
FT	143	.) (POTENTIAL)
SC	252 AA	5F10603026566BBA CRC64

Query Match: 3.38; Score 97; DB 1; Length 252;

Best Local Similarity 23.38; Pred. No. 1;

Matches	45;	Conservative	32;	Mismatches	55;	Indels	28;	Gaps	6;
---------	-----	--------------	-----	------------	-----	--------	-----	------	----

07 131 KKQVGDPELRARHSSPALTAGAGKVDNFNNGTYLVSFTLFMEGVSLILLTHPSEG-188
:
:
Db 97 KRIICGWGDY-SNKIKTVALNLVTGKIYDHGNGFSVFHRHNATGGGNSTISLVPPSKAV-155

189 -----ASALWRARNQGYDKIFKGFVNGTSHVFTECGTLNSNAELCEYLD 235

```

Db      156 EFHQEQQIFIAKASKIFNCME-WEK-VERGR-----RTSLCTHDPKICSRDH 203

```

```

QY      236 DRDQFAFYCMKPOHMPCEALTYMTTRNEY 265
          : : : : : : : : : :
DB      204 AQSSATWSCSQPFKIYGVYIAFYSTDYRLV 233

```

RESULT 13

ID	NU5M_ACACA	STANDARD;	PRT;	675 AA
----	------------	-----------	------	--------

DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE NADH-UBIQUINONE OXIDOREDUCTASE CHAIN 5 (EC 1.6.5.3).
GN ND5 OR NAD5.
OS Acanthamoeba castellanii (Amoeba).
OS Mitochondrion.
OG Eukaryota: Acanthamoebidae: Acanthamoeba.
OX NCBI_TaxID=5755;

RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ATCC 30010 / NEEF
 RX MEDLINE-95147775; Pubmed-7844833;
 RA Burger G., Plante I., Lonerjan K.W., Gray M.W.,
 RT "The Mitochondrial DNA of the amoeboid protozoan, *Acanthamoeba*
 RT *castellanii*: complete sequence, gene content and genome
 RT organization." *J. Mol. Biol.* 245:522-537(1995).
 CC -1. CATALYTIC ACTIVITY: NADH + URIDUINONE - NAD(+) + URIDUINOL

This SWISS-PROT entry is copyright. It is produced through a
collaboration between the Swiss Institute of Bioinformatics and the EMBL
European Bioinformatics Institute. There are no restrictions
on the use of this data by non-profit institutions as long as its content is
not modified and this statement is not removed. Usage by and of
this database requires a license agreement (See <http://www.isd.sib.ch>)
or send an email to license@sib.ch.

Query Match	3.38; Score 97; DB 1; Length 675;
-------------	-----------------------------------

Best Local Similarity 22.88; Pred. No. 4.1;
 Matches 45; Corresponding Mismatches 31; Indels 79; Gaps 42; 7

Matches	45;	Conservative	31;	Mismatches	19;	Indels	42;	Gaps	1
---------	-----	--------------	-----	------------	-----	--------	-----	------	---

[illegible]

RESULT 14

MSH6_ARATH STANDARD; PRT; 1324 AA.
 AC 004716;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE DNA MISMATCH REPAIR PROTEIN MSH6-1 (ATMSH6-1).
 GN MSH6-1 OR AGAA.3 OR AT4002070 OR T10M13.8.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
 OC Brassicales; Brassicaceae; Arabidopsids.
 ON NCBI_TaxID=3702;
 RX SEQUENCE FROM N.A.
 RC STRAIN=CV. LANDSBERG ERRECTA;
 RA Till S., Granat S., Parnell L., Kaplan N., Hoffman J., Lodhi M.,
 RL Johnson A.F., Dedhia N., Martienssen R., McComble W.R.;
 RL Submitted (May-1997) to the EMBL/GenBank/DBJ databases.
 RP (2)
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. COLUMBIA;
 RX MEDLINE=99444907; PubMed=10517319;
 RA Ade J., Belzile F., Philippe H., Douriaux M.P.;
 RT *Four mismatch repair paralogues coexist in Arabidopsis thaliana:
 RT AtMSH2, AtMSH3, AtMSH6-1 and AtMSH6-2.*;
 RL Mol. Gen. Genet. 262:239-249(1999).
 RN (3)
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. COLUMBIA;
 RX MEDLINE=20083488; PubMed=10617198;
 RA Mayer K.F.X., Schueller C., Wandut R., Murphy G., Volckaert G.,
 RA Pohl T., Duesterhoeft A., Stiekema W., Entlan K.-D., Terry N.,
 RA Harris B., Ansoorge W., Brandt P., Griwell L., Rieger M.,
 RA Weichselgartner M., de Simone V., Obermaler B., Macie R., Mueller M.,
 RA Kreis M., Delany M., Pulgomech P., Watson M., Schmidheini T.,
 RA Reichert B., Portetelle D., Perez-Alonso M., Boutry M., Bancroft I.,
 RA Vos P., Honzel U., Zimmermann W., Medler H., Ridley P.,
 RA Langham S.-A., McCullagh B., Bilhan L., Robben J.,
 RA Van der Schueren J., Grymonprez B., Chuang Y.-J., Vandensussche F.,
 RA Braeken M., Welljens I., Voet M., Bastiens I., Aert R., Deboer E.,
 RA Holzner E., Brandt A., Peters S., van Staveren M., Dirse W.,
 RA Moeljan P., Klein Lankhorst R., Rose M., Haut J., Koeltter P.,
 RA Bernerster S., Hempel S., Felipascu M., Lamberth S., Van den Daele H.,
 RA De Keyser A., Buyschaert C., Gielen J., Villarroel R., De Clercq R.,
 RA Van Montagu M., Rogers J., Cronin A., Quail A., Bray-Allen S.,
 RA Clark L., Doggett J., Hall S., Kay M., Leonard N., McIay K., Mayes R.,
 RA Pecterc A., Rajandream M.-A., Lyne M., Benes V., Rechmann S.,
 RA Borikova D., Bloeker H., Scharfe M., Grimm M., Loehner T.-H.,
 RA Dose S., de Haan M., Maarse A., Schaefer M., Mueller-Auer S.,
 RA Gabel C., Fuchs M., Fartman B., Granderath K., Danner D., Herzi A.,
 RA Neumann S., Argilou A., Vitale D., Liguori R., Pitarandi E.,
 RA Massenet O., Oulley F., Clabaud G., Mendenle A., Felder P.,
 RA Schenkl S., Hiller R., Schmidt W., Lecharny A., Aubourg S.,
 RA Chefdor F., Cooke R., Berger C., Montfort A., Casacuberta E.,
 RA Giddons T., Weber N., Vandenbol M., Barques M., Terol J., Torres A.,
 RA Perez-Perez A., Purnelle B., Bent E., Johnson S., Tacon D., Jesse T.,
 RA Heijnen L., Schwarz S., Scholler P., Heber S., Francis P., Bielek C.,
 RA Frishman D., Haase D., Lemcke K., Mewes H.-W., Stocker S.,
 RA Zaccaria P., Bevan M., Wilson R.K., de la Bastide M., Hadermann K.,
 RA Parnell L., Dedhia N., Gnot L., Schutz K., Huang E., Spiegel L.,
 RA Selson M., Murray J., Sheet P., Cordes M., Abu-Threideh J.,
 RA Stoneking T., Kalicki J., Graves T., Harmon G., Edwards J.,
 RA Latreille P., Courtney L., Cloud J., Abbott A., Scott K., Johnson D.,
 RA Minx P., Bentley D., Fulton B., Miller N., Greco T., Kemp K.,
 RA Kramer J., Fulton L., Mardis E., Dante M., Pepin K., Hillier L.,
 RA Nelson J., Spieth J., Ryan E., Andrews S., Geisel C., Layman D.,
 RA Du H., Ali J., Berghoff A., Jones K., Drome K., Cotton M., Joshi C.,
 RA Antonoli B., Zidanic M., Strong C., Sun H., Lamar B., Jordan C.,
 RA Ma P., Zhong J., Preston R., Yil D., Shekter M., Matero A., Shah R.,
 RA Swaby I.K., O'Shaughnessy A., Rodriguez M., Hoffman J., Till S.,
 RA Grant S., Shohdy N., Hasegawa A., Hamed A., Lodhi M., Johnson A.,

RA Chen E., Marra M., Martienssen R., McComble W.R.,
 RT *Sequence and analysis of chromosome 4 of the plant Arabidopsis
 RT thaliana.*;
 RL Nature 402:769-777(1999).
 CC -1- FUNCTION: INVOLVED IN POST-REPLICATIVE DNA-MISMATCH REPAIR (BY
 CC SIMILARITY).
 CC -1- SUBUNIT: HETERODIMER OF MSH2 AND MSH6 (THIS) (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE DNA MISMATCH REPAIR MOTIF FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/databases/>
 CC or send an email to license@sib-sib.ch).
 CC
 DR EMBL: AF001535; AB57796.1;
 DR EMBL: AJ245967; CAB5337.1;
 DR EMBL: AF001308; AAC78699.1;
 DR EMBL: AL161493; CAB80700.1;
 DR InterPro: IPR000432;
 DR InterPro: IPR002863;
 DR Pfam: PF00468; Muts_C; 1.
 DR Pfam: PF01624; Muts_N; 1.
 DR PROSITE: PS00486; DNA_MISMATCH_REPAIR_2; 1.
 DR DNA repair; ATP-binding; DNA-binding.
 KW NP_BIND; 1083
 FT NP_BIND 1083
 FT CONFLICT 316 318
 FT CONFLICT 383 383
 FT CONFLICT 852 852
 FT
 SQ SEQUENCE 1324 AA; 146797 MW; 7874340272C9799C CRC64;
 Query Match 3.38; Score 97; DB 1; Length 1324;
 Best Local Similarity 21.28; Pred. No. 10; Mismatches 126; Indels 118; Gaps 24;
 Matches 84; Conservative 68;
 QY 123 EVRDHLGQKQYGG--DPLARMSPALNAG--ASGKVMDFN-NGTVLSFLPEKVS 177
 DB 460 ETPDLEORRKRKTSKDKVVRREVCVAVTKGTLDEGLLINDASVLAALF--EGES 516
 QY 178 LSLTLHPSEGASALMARNRNGYDKLIFGRKVNQTSIVTEEGLL-NSNDELHYLD 291
 DB 517 LT-----NPI-----AEHNFQVLMVAVALUKLILLOKFD 546
 QY 237 RDEDAFYC--MKPDHM--PCEALTYMT-----TRNREVSYL-----DKNSLPH 277
 DB 547 QDSALSCLSLSEMRVEIIPAKVLSVATERIVQTRNPLVNINVLPSLHWDSKITY 405
 QY 278 RSVGVGEEMKDKRHIDVTCNKRREKLEETCOVMKPPVPGYTLQKRI--TFPCNV 433
 DB 606 --EVGIYYR-----INCP-----SSAVSEKRLCGSSSFLKML 440
 QY 334 -QIDTIKINGCIK---GKLIYLDGSTLKQMIYFPKVKVILKFTLHETGIFRKHIL 494
 DB 641 SELATEDKNGSLALSALGAIYLRQAFLEDSLRLFAK--FESLPQCDSSNVN-EKQHWL 698
 QY 389 DAEHRITQWKKHS-----YPEVFQLYSLIDH-----DYLPFI--DRLSDKN 441
 DB 699 DAAALENLEIFERNNGY---SGTLVQNLQCTASGRKLKTLWLRPLINTELKREQ 755
 QY 432 TAVITTEQOHRFRPPIDIFRRRAIGVCAIIRFLR 467
 DB 756 DAAVALRGEN---LPSVLEFRKSLSLRDPMERLIAR 788
 RESULT 15
 ID P2X2_HUMAN STANDARD; PRT; 471 AA
 AC G0UBJ9; 09Y637; 09Y638; 09Y639; 09Y640; 09Y641; 09Y642; 09Y643;
 DT 01-OCT-2000 (Rel. 40, Created)

17 01-OCT-2000 (Rel. 40, last sequence update)
 18 01-OCT-2000 (Rel. 40, last annotation update)
 19 P2X PURINOCCEPTOR 2 (P2X2) (PURINERGIC RECEPTOR).
 20 P2X2 OR P2X2.
 21 Homo sapiens (Human).
 22 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 23 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 24 NCBI_TaxID=9606;
 25
 26 [1]
 27 SEQUENCE FROM N.A. (ISOFORMS A; B; C AND D).
 28 TISSUE-Placental;
 29 PubMed=10570044;
 30 Lynch K.J., Tonna E., Nitoratos W., Kage K.L., Burgard E.C.,
 31 van Biesen T., Kowalik E.A., Jarvis M.F.;
 32 "Molecular and functional characterization of human P2X2(2)
 33 receptors";
 34 Mol. Pharmacol. 56:1171-1181(1999).
 35
 36 [2]
 37 SEQUENCE FROM N.A. (ISOFORMS A; C; H AND I).
 38 TISSUE-Prostate;
 39 Chang T.K., Kosaka A.H., Oglesby I.B., Geyer J.R., Lachnit W.G.,
 40 Ford A.P.D.W., Chang D.J.;
 41 "Cloning and molecular characterization of human P2X2 and its splice
 42 variants";
 43 Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
 44
 45 [3]
 46 SEQUENCE FROM N.A. (ISOFORMS A AND B).
 47 TISSUE-Placenta;
 48 McMahon R.A., Egan T.M., Hurley P.T., Nelson A., Rogers M., Martin F.;
 49 "Cloning of the human P2X2 receptor cDNA and multiple splice
 50 variants";
 51 Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
 52
 53 -1- FUNCTION: BINDING OF THIS LIGAND GATED ION CHANNEL TO ATP
 54 MEDIATES SYNAPTIC TRANSMISSION BETWEEN NEURONS AND FROM NEURONS TO
 55 SMOOTH MUSCLE.
 56
 57 -1- SUBUNIT: HOMO- OR HETEROPOLYMERS.
 58
 59 -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
 60
 61 -1- ALTERNATIVE PRODUCTS: 6 ISOFORMS: A (SHOWN HERE), B, C, D, H AND
 62 I; ARE PRODUCED BY ALTERNATIVE SPLICING.
 63
 64 -1- SIMILARITY: BELONGS TO THE P2X RECEPTOR FAMILY.
 65
 66
 67 This SWISS-PROT entry is copyright. It is produced through a collaboration
 68 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 69 the European Bioinformatics Institute. There are no restrictions on its
 70 use by non-profit institutions as long as its content is in no way
 71 modified and this statement is not removed. Usage by and for commercial
 72 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 73 or send an email to license@isb-sib.ch).
 74
 75
 76 EMBL: AF190822; AAF19170.1; -;
 77 EMBL: AF190823; AAF19171.1; -;
 78 EMBL: AF190824; AAF19172.1; -;
 79 EMBL: AF190825; AAF19173.1; -;
 80 EMBL: AF190826; AAF19174.1; -;
 81 EMBL: AF260426; AAF74201.1; -;
 82 EMBL: AF260427; AAF74202.1; -;
 83 EMBL: AF260428; AAF74203.1; -;
 84 EMBL: AF260429; AAF74204.1; -;
 85 EMBL: AF109387; AAD42947.1; -;
 86 EMBL: AF109388; AAD42948.1; -;
 87 MIM: 600844; -;
 88 InterPro: IPR001429; -;
 89 InterPro: IPR001045; -;
 90 Pfam: PF00864; P2X_receptor.1.
 91 PRINTS: PR01307; P2XRECEPTOR.
 92 PROSITE: PS01212; P2X2RECEPTOR.1.
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525
 526
 527
 528
 529
 530
 531
 532
 533
 534
 535
 536
 537
 538
 539
 540
 541
 542
 543
 544
 545
 546
 547
 548
 549
 550
 551
 552
 553
 554
 555
 556
 557
 558
 559
 560
 561
 562
 563
 564
 565
 566
 567
 568
 569
 570
 571
 572
 573
 574
 575
 576
 577
 578
 579
 580
 581
 582
 583
 584
 585
 586
 587
 588
 589
 590
 591
 592
 593
 594
 595
 596
 597
 598
 599
 600
 601
 602
 603
 604
 605
 606
 607
 608
 609
 610
 611
 612
 613
 614
 615
 616
 617
 618
 619
 620
 621
 622
 623
 624
 625
 626
 627
 628
 629
 630
 631
 632
 633
 634
 635
 636
 637
 638
 639
 640
 641
 642
 643
 644
 645
 646
 647
 648
 649
 650
 651
 652
 653
 654
 655
 656
 657
 658
 659
 660
 661
 662
 663
 664
 665
 666
 667
 668
 669
 670
 671
 672
 673
 674
 675
 676
 677
 678
 679
 680
 681
 682
 683
 684
 685
 686
 687
 688
 689
 690
 691
 692
 693
 694
 695
 696
 697
 698
 699
 700
 701
 702
 703
 704
 705
 706
 707
 708
 709
 710
 711
 712
 713
 714
 715
 716
 717
 718
 719
 720
 721
 722
 723
 724
 725
 726
 727
 728
 729
 730
 731
 732
 733
 734
 735
 736
 737
 738
 739
 740
 741
 742
 743
 744
 745
 746
 747
 748
 749
 750
 751
 752
 753
 754
 755
 756
 757
 758
 759
 760
 761
 762
 763
 764
 765
 766
 767
 768
 769
 770
 771
 772
 773
 774
 775
 776
 777
 778
 779
 780
 781
 782
 783
 784
 785
 786
 787
 788
 789
 790
 791
 792
 793
 794
 795
 796
 797
 798
 799
 800
 801
 802
 803
 804
 805
 806
 807
 808
 809
 810
 811
 812
 813
 814
 815
 816
 817
 818
 819
 820
 821
 822
 823
 824
 825
 826
 827
 828
 829
 830
 831
 832
 833
 834
 835
 836
 837
 838
 839
 840
 841
 842
 843
 844
 845
 846
 847
 848
 849
 850
 851
 852
 853
 854
 855
 856
 857
 858
 859
 860
 861
 862
 863
 864
 865
 866
 867
 868
 869
 870
 871
 872
 873
 874
 875
 876
 877
 878
 879
 880
 881
 882
 883
 884
 885
 886
 887
 888
 889
 890
 891
 892
 893
 894
 895
 896
 897
 898
 899
 900
 901
 902
 903
 904
 905
 906
 907
 908
 909
 910
 911
 912
 913
 914
 915
 916
 917
 918
 919
 920
 921
 922
 923
 924
 925
 926
 927
 928
 929
 930
 931
 932
 933
 934
 935
 936
 937
 938
 939
 940
 941
 942
 943
 944
 945
 946
 947
 948
 949
 950
 951
 952
 953
 954
 955
 956
 957
 958
 959
 960
 961
 962
 963
 964
 965
 966
 967
 968
 969
 970
 971
 972
 973
 974
 975
 976
 977
 978
 979
 980
 981
 982
 983
 984
 985
 986
 987
 988
 989
 990
 991
 992
 993
 994
 995
 996
 997
 998
 999
 1000

FT DOMAIN 359 471 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 320 333 PORE-FORMING MOTIF (POTENTIAL).
 FT CARBOHYD 133 133 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 194 194 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 310 310 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT VARSPPLIC 36 127 MISSING (IN ISOFORM H).
 FT VARSPPLIC 38 152 NRRGVYVRAVOLLLTFYVYVYFYOKSYOPESESESSI
 FT ITKKKITTSEKRWVDEVEYVYKPEEGGSFYITITVETATHS
 FT OTQGTCESTVYVHATCLSDDCVAGELMDLON -> ITHRA
 FT EKLDEGDEPRELHHGQGDHVAHRAOSVGRGVGEAPR
 FT (IN ISOFORM I).
 FT VARSPPLIC 104 127 MISSING (IN ISOFORM C).
 FT VARSPPLIC 354 354 V -> VWRPMLGPGCGSGSTPRLHTGLCPQ (IN
 FT VARSPPLIC 381 447 ISOFORM D).
 FT CONFLICT 1 14 MISSING (IN ISOFORM B).
 FT MAAPQRYRAGATA -> MV (IN REF. 3).
 SO SEQUENCE 471 AA: 51754 MW: 84061DA136E420 CRC64;

Query Match 3.3%; Score 96.5; DB 1; Length 471;
 Best Local Similarity 21.4%; Pred. No. 2.7;
 Matches 88; Conservative 46; Mismatches 120; Indels 157; Gaps 23;

QY 11 LILISPSVYVMMFIISONTFKLMSALNISVHWNNSAKSLFPKTSLI-PLKPLTET 69
 DB 50 LILILYF--VWYFYVOKSYOE-----SETGPESSITVYKGIITS 88
 QY 70 ELRIKEIIEKLDQIPRP-----FTVNTTTSATHT-----ATLNPRTQYR 114
 DB 89 EKHVMDVEEYK--PREGGSVFISITTEVHTSQTGTGCPESIVHNACTLSDAD--CV 143
 QY 115 GQGLDILLEVDHIGRKQYGDPLRA-----KSSPALTAGAS-----KYVDFNN 161
 DB 144 AGELEML-----GNGLRTRCVPYVQGSKTCEVFQWCFVEDGASVQF-L 188
 QY 162 GTIYVFTLFMEGVSLDLLIPSEGASALMARNOGYDK-----IIFK-GK 208
 DB 189 GTVAPNFTILIKN--SIHYKPFHFKGNIA---DRIDGYLKCTHESADLYCPFKLGF 243
 QY 209 FVNGSVHFT-----CGILNNAALCEYLDLDDQDAFYC-----MPPQMPCE 253
 DB 244 IYKAGESFTLAHKGVGIVGIINWDCD-----DLPASCNPKYSFRLLDPKHPAS 296
 QY 254 A-----LTVMTRN-----REVSYLTDKENSILFHSKGVG-E 284
 DB 297 SGVNFPAKYKIKGTTRTLKAVGIRIDYVHGQAGKFSPLITLMLATLTSVGVS 356
 QY 285 MKDKRKHIDVYCNCK--REKTEFCO-----VGKKPVPVG 317
 DB 357 FLGDWILITFNKKRYSHKFKEDVCTPSHPSGMPVTLARVLGQAPPEFG 407

Search completed: October 12, 2001, 16:05:44
 Job time: 187 sec

09Y26 PRELIMINARY: PRT: 860 AA.
 ID 09Y26:
 AC 09Y26:
 DT 01-MAY-1999 (TREMblrel. 10, Created)
 DT 01-MAY-1999 (TREMblrel. 10, Last sequence update)
 DT 01-MAR-2001 (TREMblrel. 16, Last annotation update)
 DE ORF MSV067 PUTATIVE mRNA SPINNING ENZYME LARGE SUBUNIT HOMOLOG
 DE (VACCINIA DLR), SIMILAR TO SW:P20979.
 GN MSV067.
 OS Melanoplus sanguinipes entomopoxvirus (MSEPV).
 OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Entomopoxvirinae;
 OC Entomopoxvirus B.
 OX NCBI_TaxId=63191;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=TUCSON.
 RX MEDLINE=9102612; PubMed=9847359;
 RA Alfonso C.L., Tulman E.R., Lu Z., Oma E., Kutish G.F., Rock D.L.;
 RT "The genome of Melanoplus sanguinipes entomopoxvirus";
 RL J. Virol. 73:533-552(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=TUCSON.
 RA Alfonso C.L., Tulman E.R., Lu Z., Oma E., Kutish G.F., Rock D.L.;
 RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF063866; AAC97815.1;
 SO SEQUENCE 860 AA; 101503 MW; 30420ACBE1037692 CRC64;

Query Match 4.1%; Score 119.5; DB 14; Length 860;
 Best Local Similarity 18.3%; Pred. No. 0.25;
 Matches 132; Conservative 96; Mismatches 207; Indels 285; Gaps 31;

QY 4 NTML---OKTLLILISFVYTMPIISONT---KLMSAL-----NLSTSVHYHNSAK 52
 DB 126 NMNIIPKKILMSVNTIHLISSTIYINGVREFLKTGLADANNKMLQYTLN--- 182
 QY 53 SLFPTSLIPKPLTELEIKELIKDOQIRPRTF-----VN-----TTTSA 98
 DB 183 -----VLTQKQTSYIEI-ELINKLSODEFYRITAFRYISINNINNFALSTKNI 234
 QY 99 THSTATILNPRDTCRGDQLDILLEVRDHLGOKKQYGDFLRAKSSPALTACAGKYWD 158
 DB 235 SPSTVTFMLPFDLSL-----YLMDSKY-----FLTSKIDGEVQ 268
 QY 159 FNGGTLYVSFTLMEGQVSLSLILHPSBQASALMAROGYDKITFGKFNVTSH--- 215
 DB 269 FTVKNGIDILVY-----VFTECGLT-----LNSNAELCEYLDRODEAFYCMKPOHMP 252
 QY 216 -----VFTECGLT-----LNSNAELCEYLDRODEAFYCMKPOHMP 252
 DB 299 KQFGEYKKIDNVKTYIPFVFEIYSTDSKQYNTREMIKFNNDINKNEFDITFKHKPI 358
 QY 253 -----EALTYMTRRREVS---YLDKENSFLHRSKY----- 281
 DB 359 IPIEKDIYNEALAYKYSIDEKITGVLDDKESNIDYVTLKIDNTVDVIALSTHKSPIKI 418
 QY 282 ---GVEM-----MKDKHIDVTONKREKIEGTQVCGKRPVPGYTLOGKAITTCNQV 333
 DB 419 HKDGIYTFNLVYNDKKN-----ITELLVEMK-----SNDLIYNDI 457
 QY 334 QLDTIKINGCLGKLYLLGDSLTLMQIYYPKVV-----KTLKFDLDEGTIEFKHLL 368
 DB 458 NMLKFKNIHNRGKNILYAPLCCLVIESFLKSKIVINIMDKTKTKFKSKYVYNSLK-VIL 516
 QY 389 DAERHTQIQMKKHSYFETPOLYSLDHYITREIDRLSGDKNTAIV-----IT 437
 DB 517 DS-----KTFHKKYVPLDTVG-IDYLYTMYFKINMEIYDKNLLLNONEVKKYKPEIE 567
 QY 438 FGHFRFPIDIFIRRAIGVOKAIERLFRSPATVILKTEIRHNIETFRGPGFYI 497
 DB 568 NNOOORP-PLNIF-----TNLIKTEAISAAKSLCAELIPNHYVLSIDIGRGDINKYI 619

QY 498 HYILM-----KDI-----FKELNVCILN----- 679
 DB 620 YIGISGMLGTDPDISALSEADENYNSLQTRKRVSTYFKSSLNLSILDDHYLKKVSY 679
 QY 516 -----ANDMTI-AYGDTIHPPHVI-----GNCI-NM-LN 644
 DB 680 MTOOKIKFYGLIEWQMAIHYSVVDTKKILNLIKLNSTDGTKEVITTCNCKIKILLIN 739
 RESULT 3
 ID 09ESN6 PRELIMINARY: PRT: 744 AA.
 AC 09ESN6:
 DT 01-MAR-2001 (TREMblrel. 16, Created)
 DT 01-MAR-2001 (TREMblrel. 16, Last sequence update)
 DT 01-MAR-2001 (TREMblrel. 16, Last annotation update)
 DE NEURAL ACTIVITY-RELATED RING FINGER PROTEIN.
 GN NARF.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxId=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ICR MICE;
 RA Ohkawa N.;
 RT "cDNA sequence encoding Neural Activity-Related Ring Finger Protein
 (NARF).";
 RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AB043550; BAB17634.1;
 SO SEQUENCE 744 AA; 81445 MW; 80346DA2E9E7E9C3E744;

Query Match 4.0%; Score 117; DB 11; Length 744;
 Best Local Similarity 28.5%; Pred. No. 0.33;
 Matches 41; Conservative 23; Mismatches 50; Indels 70; Gaps 25
 QY 67 TETETRI--KEIEKL---DOQIRPRP-----FTVNTTSA 98
 DB 266 TETEVLLVKQMSKELNLADDPPLHPRENQDLFIETEGCKKSTHNGITLITNNAV 425
 QY 99 THSTATILNPRDTCRGDQLDILLEVRDHLGOKKQYGDFLRAKSSPALTACAGKYWD 158
 DB 326 SEYVATGEGIKQITII-GQPMSTVITTDKNGELCKTQNAVILALSTPDSV-ALGELL 364
 QY 159 FNGGTLYVSFTLMEGQVSLSLILHPSBQASALMAROGYDKITFGKFNVTSH--- 215
 DB 384 NKNGTYEFLLTYQKRGDFTLSURL 407
 RESULT 4
 ID 060272 PRELIMINARY: PRT: 792 AA.
 AC 060272:
 DT 01-AUG-1998 (TREMblrel. 07, Created)
 DT 01-AUG-1998 (TREMblrel. 07, Last sequence update)
 DT 01-MAR-2001 (TREMblrel. 16, Last annotation update)
 DE KIA0517 PROTEIN (FRAGMENT).
 GN KIA0517.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxId=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=BRAIN;
 RX MEDLINE=98290545; PubMed=9628561;
 RA Nagase T., Ishikawa K., Miyajima N., Tanaka A., Kotani H., Nomura N.,
 RA Ohara O.;
 RT "Prediction of the coding sequences of unidentified human genes. IX.
 RT The complete sequences of 100 new cDNA clones from brain which can
 RT code for large proteins in vitro.";
 RL DNA Res. 5:31-39(1998).


```

QY 42 ISVHWNNKASLFPKTSUPLKP-LTEELRIKEIEKLDQDIPRPTHTVTTTSATH 100
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 230 LEAHFGKILREELGIKSSI--TRPILRLNVLRRKRLVVLVDVCKPL----- 274
QY 101 SPATILNPRDTCGRD-----QDILILEVRD-----HIGORKYGDGF 138
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 275 DAESEFLGCFMFCFOSLIITTSRDKOVSGICRVDOLEYEVLNLEEDALDIFSCAFGKEI 334
QY 139 LRAHSSPALTAGASGVADFNNGTYLVSTFLFMEGVOSTS---LLLIHPSDG---ASA 191
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 335 IHESLQK-----LSKRVIVDANGNPL--ALIFGCHSRKNRPPIEIAFPVKRYLAHE 385
QY 192 LMRARNGCXKI-----IKKGFVNGISVHFTGCG---LTISSNAELC---231
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 386 IHDVAKSTSDSSNEKNIFLDIACLRGENVDVHLLDECGCFPRVEINVLKELCVS 445
QY 232 -----EYLDOR-----DOEAFYC-----244
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 446 MAEGRVVHNLIOISGRKIIINGKRSRLMKPLLIKRYLEDRVLGSEDEIAIFLDPAL 505
QY 245 ---MKP-----OHMPCEALTY- 257
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 506 SFVNPAPAFENMNTRLYKICSSNPGNHVHLLPRGVKSLPEELRLHWEHFPPLSLPEOD 565
QY 258 MTRNREVSYLTDKENSLEFHSKVGVEYMKDKRHIDVTNCKRKEIETCOVGMKPPYPG 317
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 566 FTRNLVILNMC---YSKLQRLMEGTKEIGMLKRIMLCHSOLVIGIE-LQIALMEV-- 619
QY 318 GYTLQGWITTFGNVO-----LDTIKINGCLKGLIYLDSDTLROWIYTPKY 367
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 620 -IDLQ-----CARLQRFATGHPHLLVNLVLSGIC-----IKSPEV 657
QY 368 VKTIKFFPDHETGIFRKHLLDAERHTOIOMKHSYPTVFO-----LYSLIDHDYPR 421
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 658 PPNIEELYLKQNGI-----RSIPVTFVSPQDNSFYDHMKHKNLR 698
QY 422 EIDRLSGDKNTAVITFGQHRFPID--IFIRALIOVAIRLRLSPATK----- 472
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 699 EV---SSDSOSLSIMVILDNKVLVLSGLELIEDIGIPNKLKYLGTGAIKELPSLMH 755
QY 473 ----VIKTEIREMH 484
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 756 LSELVVLDELCKRLH 771
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 7
ID 004408 PRELIMINARY: PRT: 801 AA.
AC 004408;
DT 01-JUL-1997 (TREMBLrel. 04, Created)
DT 01-JUL-1997 (TREMBLrel. 04, Last sequence update)
DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
DE ENT-KAURENE SYNTHASE A.
GN LS.
OS Plasm sativum (Garden pea).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
OC Fabales; Fabaceae; Papilionoideae; Plsum.
OX NCBI_TaxID=3888;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. TORSDAG;
RA MEDLINE=97260958; PubMed=9107034;
RA Alt-All T., Swain S.M., Reid J.B., Sun T.P., Kamiya Y.,
RT "The 1S locus of pea encodes the gibberellin biosynthesis enzyme ent-
RT kaurene synthase A".
RL Plant J. 11:443-454(1997).
DR EMBL: U63652; AAB58822.1;
DR Mendei: 16494; PIsa:1532;16494.
DR InterPro: IPR001906;
DR Pfam: PF01397; Terpene_synth; 1.
SQ SEQUENCE 801 AA: 92717 MW: 50884ADE00366844 CMC64;

```

```

Query Match 3.5%; Score 103; Db 10; Length 801;
Best Local Similarity 19.8%; Pred. No. 5.5;
Matches 124; Conservative 91; Mismatches 107; Indels 214; Gaps 0;

QY 2 SSNTMOKTLLISFVSVTW-----MFIISNFKLSALNLSIVHWNNKSA 61
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 154 SAHRIINTLACVILR--SNMHSKQKGMIF--FRNLKLENNHEHM 120----- 404
QY 52 KSLPFTSLIP---LKPDL--TETELRIKEIEKLDQDIP--PRFPVHNTTSATHST 104
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 206 DVAFP--SLIEGARGIKPLMCPNDSP-LKNIFERDKEIKRIKELIMKVP?----- 256
QY 103 ATILNPRDTCGRDQDIL-LVDRHLDGRKRYGDFIRAMSSPALTAGASGVAMFEN 141
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 256 -TLHSLEGMSGLDKQLKLOSD-----GSFL---SPSSTAFALMTKQDNC 301
QY 162 GTYLVSTFLFMEGVOSTSLLLIHPSGASALW--RAVNGYIKIIFKGF--VNGISH 215
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 302 LKYLNVVVKFNGVVP--NVYPLDFEHWVGRLEKLGISK--FFHHEIKLCNVYSK 309
QY 216 VTEEGCL--TNSNAELCEYLDRODEA-----PYMKP 247
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 357 INSEKICARNSVQ-----VIDDTAAFFELRLGHQVSAHVFKHPEHNDFFCPAG 410
QY 248 QHMPG-EALTYMTRNREVSYLTDKENSLEFHSKVGVEYMKDKRH----- 241
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 411 Q---CTQAVTGMYNLFASQVLPFGCKILEHAKHFSKAVLKEKREANLILKMLIMNLP 407
QY 292 -----IDVTGNCKRKEIETG---QVGMKPPYGVGTLOGKAITTEGQVVLDTIKNG 342
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 468 EEVGYALDMPVAMLDRIETRFIDYGAESVAMIGTGLRMAVYNNVNYIELKLYNN 527
QY 343 CLKGLIYLDSDTLRQWIYFPKVVTKLFED-LHETGIFRKHLLTAERHVLQMKKH 401
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 528 QQAQHL-----EW-----NVIGTWLESRLGFGSKULL----- 560
QY 402 SYPVTFQYSLLIDHYIPREIDRLSGOKNTAVITFGQHRFPIDIFIRKAGVKA 461
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 561 AYFLATGSIPE-----PERSHERLAMAKTIALLET----- 541
QY 462 ERLFLSPATKVIITKTEINDEMIETEPGFHGYIHYLMKTIK?-----LN 510
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 592 -----KCYVRNEDLRKDF--AKKFNHIDVRSYARRHMKKTEHEVESLFAF 639
QY 511 VGLIDAMDTIAYGDTIHPPDHVIG 536
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 640 IGEI-SMDVRLSYG-----HEIG 656
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 8
ID 09JG58 PRELIMINARY: PRT: 765 AA.
AC 09JG58;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE PORFL.
OS TT virus.
OC viruses; ssDNA viruses; unclassified ssDNA viruses.
OX NCBI_TaxID=68887;
RN [1]
RP SEQUENCE FROM N.A.
RA Tanaka Y., Mizokami M., Orto E., Ohno T., Nakano T., Hayashi K.,
RT "Identification of a 23kDa protein encoded by putative open reading
RT frame 2 of TT virus (TTV) genotype I different from the (T)iv
RT genotypes".
RL Submitted (JUL-1999) to the EMBL/Genbank/DDBJ databases.
DR EMBL: AB030487; BAA90406.1;
DR F000000000;
SQ SEQUENCE 765 AA: 90225 MW: E845A26B69D9D707 GKR64;

```

Query Match 3.5%; Score 100.5; DB 14; Length 765;

Best Local Similarity 20.2%; Pred. No. 8.3;
Matches 94; Conservative 74; Mismatches 190; Indels 107; Gaps 23;

```

QY 14 LLSFVVTWM-----FIISQNTKLMSALNLSIVHYMNSAKSLFPR-TSLPLKPLTE 68
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 264 VVTFQVLRSMYNDALSLIFSNFEO-----DONAQKLYNEISSYPYNTTE 309
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 69 TELRIKEIIFKIDQ-QIPRPF--THVNTTSAHSTATILNP-----RDYCHGDQDLIL 121
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 10 TTAQIKREVENTEKISTTPNPMQSNVYNTITTTAQSITTTTPTTSDSYWGYATYKKA 369
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 122 LEVRDHLGCRKUYGDFLARSPSPALTAQASGVNDENNGYVLSFTLFMEQVSLSL 181
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 170 I-TKVLPLAAKLY--ETQTKNLSPTFGQS--EYLEYHGILY----- 407
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 182 LHPSEGASALMRANOGYDKIIFGKGFVNGTSHVTEGGLTNSNAELCEYL--DDRD 238
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 408 -----SSITWLSAGRSYFET--KGAVTDICYNPTDNG--EGNMLMIDWLSKQDSRY 454
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 239 OFAFYCKPQJHMFCEALTY-----MTYNNREVSILTUKENSLFHNSKGVEMKDKKH 291
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 455 DKARSKCLIEKLPMMAAVYGAECYCAKATGDSNI-----DMNARYVMGCPTYVQIMID-- 507
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 292 IDVTNCKNREKIEETCOVMKRPVPG-----YTLOGKWITTFPCNOVQDLTIKINCLEK 346
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 508 ---TSDILRGFIYSPNFG-KCKMNGGNGOVPIMBAKVP--CLFHQKEVLEAIG-QSG 560
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 347 KLIY-----LLGDSITLKNQIY----YFRKYVKTIKFIDLHEGIGIRKHLIDAEHTQ 395
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 561 PFAVHSDQKAVVLGLKFRPHMIMGPNPFOVVRNCKDQGSTPRK-----PRSVQ 613
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 396 IQMKKHSYFVFOLYSLIDHYIPREIDRLSGDKNTAIVITFQ 440
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 614 IIDPKYNTPELTIHAMDPRGFRGFKALIKMQQPTDAELLPGR 658
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

```

```

QY 128 LGOKROYGDFLARMSPALTAQASGVNDENNGYVLSFTLFMEQVSLSLIHPSE 187
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 378 --PRQAAATLYSKATKTLTGLNFTTDDHTEIYEHGILY----- 411
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 188 GASALMRANOGYDKIIFGKGFVNGTSHVTEGGLTNSNAELCEYLDNR---DOAFY 243
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 412 --SSITWLSAGRSYFET--TGAVTDICYNPTDNG--EGNMLMIDWLSKNNMYDKVSK 464
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 244 CKRPQHMFCEALTY-----MTYNNREVSILTUKENSLFHNSKGVEMKDKKH 287
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 465 CU-TSDILPLAAAVGVFEQAKSGDGNHMMARLLINSPTDQILY----- 511
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 288 DRKHIDVTNCKNREKIEETCOVMKRP-----PYPGSYT-----LOGKWITTFPCNOV 334
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 512 ---HTDPFK-----GVFYSLNFNGMKMGSSNVPFRMAKWPFLFHQOE 555
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 335 L-----DTIKINGCLKGLIYLLDSTIKQMIYFPPKVKIKLAFPLDHTG 380
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 556 VLEALQSGFPAYHSDIKKVSIGMKYFPMKMGNNVROVVRNP-----CKETHSG 608
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 381 IFRKHLIDAEHTQIQMKKHSYFVFOLYSLIDHYIPREIDRLSGDKNTAIVITFQ 440
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 609 -----NRPVRSIQIVDPKYNSELTFTHTMDFRGLGPKAIOPMQOQPTTIDIFSAG- 660
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 441 HFRP 444
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 661 HKRP 664
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

```

Query Match 3.5%; Score 100.5; DB 14; Length 770;
Best Local Similarity 19.4%; Pred. No. 8.4;
Matches 94; Conservative 66; Mismatches 171; Indels 153; Gaps 21;

Query Match 3.4%; Score 100; DB 2; Length 527;
Best Local Similarity 20.0%; Pred. No. 5.3;
Matches 92; Conservative 68; Mismatches 158; Indels 142; Gaps 25;

```

QY 30 NFKTLMAL--NLSIVHYMNSAKSLFPRKPLTELRIKRIIEKLDQGIIPR 87
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 265 NFOVLOSMYQNTSILPDKSERKELTQIDVYIPYNTTQITQALAKFID-AGNTPTMT 323
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 88 PFT-----HVNT--TISA-----THSTATILNPRODTYRG-----DQDILILEVRDH 127
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 124 PATTWASYINTTKFTTAATTIYTPGTITTYTMLCDSWYRGVYDKIKNL----- 377
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

```

```

QY 65 PLETELRI--KEIIEKLDQOI--PPR-----PFTVNTTTSATHTATILNPRDYCR- 114
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 2 PLLSLRLIIONSPIRKYNNQVDTSPKIGICVPLFSIHQNSCGIGERLDLIMIDWCTL 61
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 115 -GDQDILILEVRDHLGCRKROYGDFLARMSPALTAQASGVNDENNGYVLSFTLFME 173
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 62 CGFQILQILPINDGSCSSPYNS-----ISSISL-----NPLHLSISALPYK 103
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 174 GQVSLSLIHPSEGASALMRANOGYDKII-FKGKGFVNGTSHVTEEC-GLTNSNAELC 231
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

```

```

Db      104 EEVSSSRKLIQEMORLSOLSOY-----NNEKVIPIKRAFF-----KEYFRVCKSKNLI/NHPDFC 157b
Oy      232 EYLDNRD-----EALYCKRPHQMBPCALYMYTNRREVSYLTQENSL-----FRRS 279b
Db      158 DFC-EREVYMLHAYALFCSIRHLNLYPIHNMSTYYTDLSTISHEHETFAKIDFYSTLQ 216b
Oy      280 KVGYSNMD-RKHLIDVTGNCKREKI-----EELCOY-----GKRPVPGGYTLQ 325b
Db      217 YLCEFOMKQVRKHAHDKGLLIGDPIILISKDSQDMFYRKYFSSSESGSP-PDFYNAE 275b
Oy      323 G-WITTFQNOVOLDITKINGCLGKLIYLLGDSILRQMIYFPKVVKTLK-----FEDL- 376b
Db      276 GQNNMLPIYNN-----KTLRQDAYHMMK--ERLRYAENFYELY 311b
Oy      377 ---HEETGFKHLLDAENHTQIQMKHSPVYTFOLSLDH-----DYRPEID 424a
Db      312 RLIDHYVIGFR-----FWVDELGRGRREPQDDPYDITDOSTD 347b
Oy      425 RUSGKNTALYITFGHFPPFIDLFRRALGYOKAIERL 464
Db      348 ILSHLLKASMLPIGDGLTTPVD-----VKQALESL 379

```

RESULT	11		
Q9RLP9	Q9RLP9	PRELIMINARY;	PRT; 654 AA.
AC	Q9RLP9;		
DT	01-MAY-2000 (TREMBLrel. 13, Created)		
DT	01-MAY-2000 (TREMBLrel. 13, Last sequence update)		
DT	01-MAY-2000 (TREMBLrel. 13, Last annotation update)		
DE	P75 PROTEIN PRECURSOR.		
GN	P75.		
OS	Mycoplasma hominis.		
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;		
OC	Mycoplasmataceae; Mycoplasma.		
CC	NCBI_TaxID=2098;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=183;		
RA	Myrind T., Birkeland S., Christiansen G.;		
RT	"Cloning, DNA sequencing and restriction enzyme analysis of the p75/51 gene indicates the presence of a different type of surface protein variability in Mycoplasma hominis.";		
RL	Submitted (JUL-1999) to the EMBL/Genbank/DBJ databases.		
DR	EMBL: AJ243903; CAB02240.1; -		
KW	Signal.		
FT	SEQUENCE	1	27
SO	SEQUENCE	654 AA;	73607 MW; 38E7C46FF5D44091 CRC64;
			POTENTIAL.

Query Match 3.4%; Score 99; DB 2; Length 654;
Best Local Similarity 18.5%; Pred. No. 8.9;
Matches 103; conservative 98; Mismatches 188; Indels 168; Gaps 28;

```

0Y      2  SENTLQKTLJLJISFVYMWI-FIIISOFRKLMSALNISVHWN--NSAKSLPFKT 58
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db     122  SONDLKK-----LVELEITTKLMMLSNESH-----NITKEAKNLLMNALEKS 173
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
0Y      59  S-----LIPKPLTELETRAKELIEKLDQIIPRPPTHVNTTSATSIINPRDYCR 114
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db     174  NQOEYKNOIKRVNETINKTEEISINLEKGRKSEFTTKE-----LREKLTIVYL 222
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
0Y     115  GDOL-----DILLEVRHGLGORKOYGGFLRARRSPALTAGACGVDFNNGYIVSE 168
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db     223  DENTSEFAYMDLAEIKELIQOENV-----SIVSNKEIKSAISOVD----- 266
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
0Y     169  TEFMEGVOSTL--LIH-----PSEBASLMARNOQYDIILKRFVN-----GTS 214
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db     267  -TFWYANIKKSIIIDKFLYLENKRSLDVKFLAHDELSQLWTSIFENIETTERBL 375
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
0Y     215  HYFIEGGLTNSNAELCEYLDROEAFYOMKPOHMPCEALTIVYTTNRREVSYLTDKENS 274
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :

```

```

Db 326 NKIKTNLTKISNVEKQNIIE-----VNNTNS 464
QY 275 LPHRSKVEVEMKDRKHT---DVNCKKRRKEIEIICQVMKRPVPSYSLCKNIIITP 78 441
Db 355 LNKNNNNNNKNDQIYELSGDIEOLKRLQIIT-----NNVHIVKRIINLN 30
QY 332 QVODLTIKNGCKGLIYL--LGDSTLPMIYPPKPVKTLKFFLHETGIRKKHLLD 76
Db 403 KA--NTIKNEGSPKAKAILLETLRD-TISELFELYLQDLHLHLEFFIKNNIDIK--LLP 40
QY 390 AERHGIQOMKHHSPVPTQLYSLIDDDIYPRKIDKLSLCKNT---AVITGQJHRIPT 44
Db 457 ASDSKRI-----NDYL-NKIKKIT-ENNSCTPQSETTPELYISAF 444
QY 447 IDIFIRRAIGVAKIERLE-----LRSPATVILK-TENIRSMHIEETPRGPH--- 404
Db 495 -----KAOEIEPKIKKSDKADVNIEKLTNNLKE--IKREKKNDVNLVVT 540
QY 495 GYIYHILKIDFEDLVV 511
Db 541 SYLKEKIEETITKNESI 557

```

RESULT	12	
ID	012043	
AC	012043	PRELIMINARY:
DT	01-NOV-1996	PT: 749 AA.
DT	01-NOV-1996	(TREMBLrel. 01. Created)
DT	01-MAR-2001	(TREMBLrel. 16. Last sequence update)
DE	CHROMOSOME XV READING FRAME ORF YOR0801C (YOR2y-32 PROTEIN).	
GN	YOR2964C OR YOR2y-32 OR YOR081C.	
OS	Saccharomyces cerevisiae (Baker's yeast).	
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;	
OC	Saccharomycetales; Saccharomycetaceae; Saccharomycetes.	
OX	NCBI_TaxID=4932:	
RN	[1]	
RP	SEQUENCE OF 434-749 FROM N.A.	
RA	Bohn C., Bolotin-Fukuhara M., Daignan-Fornier B., Dang D.V.,	
RA	Valens M.;	
RL	Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.	
RN	[2]	
RP	SEQUENCE FROM N.A.	
RA	Voss H., Benes V., Rechmann S., Teodoru C., Schwager C., Paces V.	
RA	Assorge M.;	
RL	Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.	
RN	[3]	
RP	SEQUENCE FROM N.A.	
RA	MIPS.	
RL	Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.	

RP SEQUENCE FROM N.A.
 RP Benes V., Andrade M.A., Rechmann S., Teodoru C., Banerji A.,
 RA Sander C., Valencia A., Anisage M., Voss H.;
 RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
 RN (5)
 RN SEQUENCE OF 435-749 FROM N.A.
 RX MEDLINE=9729235; PubMed=5133743;
 RA Valens M., Bohm C., Daighan-Fornier B., Dang V., Holotini-Fukuhara M.;
 RT "The sequence of a 54.7 kb fragment of yeast chromosome XV reveals the
 RT presence of two tRNAs and 24 new open reading frames.";
 RL Yeast 13:379-390(1997).
 DR EMBL: Z74089; CAA99274.1; -;
 DR EMBL: X94335; CAA64004.1; -;
 DR EMBL: Z70678; CAA34566.1; -;
 DR SGD: S0005607; YOR081C.
 SQ SEQUENCE 749 AA: 84715 MW: F365DBA97BD6A58b CR664.

Query Match 3.48; Score 98; DB 3; Length 749;
Best Local Similarity 20.48; Pred. No. 13;
Matches 120; Conservative 70; Mismatches 224; Indels 174; Gaps 21.

```

OY 5 TMLQKTLILISFVSVMIFLISQNTKLSALNLSIVHVNNSAKSLFPTSLIPK 64
DB 37 TLSRKOLLI-----SDLSQKHAISYQOMDNLASRLDITGLSEMK 78
OY 65 PLTELEIRIKEIIEKLDQOIPPFPHVTNTTSATHTATILNPDYCRGQDLDILEV 124
DB 79 TIDESSLYNKLILODITRM-----RLRTTHD-YHRLLYLFTKVRNLGNMNVNLYR 132
OY 125 RDHIGORQYGGDFLRANSSPALTAGASQVMDNNTYLVSTLFWESQVLSLLIH 184
DB 134 HSHTG-TKQIINDYLEE--SQAVLTA-----LIR 158
OY 185 PSE-----GASALMRRNOCIDKTIIFKQVNTSVTEGGLTNSNAELCEYLD 237
DB 159 OSNMNDHYLLGLQOFR-RNIGRTALVLSGSGTFGLFIVGLAAL--FESDLMKVI 214
OY 218 EQ-----EAFYCMKPOHMPCEALTYMT--TRNREYSLTDKNSLFRSKVGVEMKDKKH 291
DB 215 SSAGATVASIFCV--HTQOIPSLILNVLNMEFIEND-DNS-----253
OY 292 IDVTNCRKREKIEETQCVGKMPYVPGGYTLQCKMTTTCQVOLDTIKINGLCKKLYL 351
DB 254 -KSPENLLIKISFCQNG-----TWENNQPLINT-----MLSF 286
OY 352 LSDSTLRQ-----WIYEPKVKVTKLFEDL-----HETGIFKKH 385
DB 287 LGMLTFREAYNKTGKLINTVSPASIEQPKLNNLAPNLWMSAVCSGLGVEPST 346
OY 386 ILDAERHTQIOMKHSFPVTFQLSL-----IDHDYIPREIDLSGDKNTAVITFCG 439
DB 347 PLFEKDPHT--GKKEKGNATNLHLSNMKMFMDGSDND--MPSLSEMRNVDHIIAQ 400
OY 440 QHFRPPIDIFPRAIG--VQKAIERLPLKSPATKV--IKTENIEMHIEFERGQPHG 495
DB 401 VNIHVPFLKFSNTQVGEIEKEITARF-RNQVTKIKFPSDETINFDILKEL--EFH 457
OY 496 YIHVLMKQDFKDLNGLIAMDMITAVGTIIPRPHVGNQINMEL 543
DB 458 YL MTKLKHFLQOYSG-----NTTILPDLMSVQCFHEVLKNSQFL 499

RESULT 13
OY 059046 PRELIMINARY: PRT: 761 AA.
AC 059046:
DT 01-JUN-1998 (TREMblrel. 06, Created)
DT 01-JUN-1998 (TREMblrel. 06, Last sequence update)
DT 01-MAR-2001 (TREMblrel. 16, Last annotation update)
DE HYPOTHEICAL PROTEIN M01652.
GN M01652.
OS Methanococcus jannaschii.
OC Archaea: Euryarchaeota: Methanococcales: Methanococcaceae:
DB Methanococcus.
OY MCBL_TaxID=2190:
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DSM 2661.
RX MEDLINE=96337999; PubMed=8688087;
KA Ruit C.J., White O., Olsen G.G., Zhou L., Fleischmann R.D.,
KA Kurland V.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
KA Overbeek R., Kirkness E.F., Weissflog K.G., Merrick J.M., Glodek A.,
KA Scott J.L., Goodenough N.S.M., Weisman J.F., Fuhmann J.L., Nguyen D.,
KA Uiterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
KA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
KA Kleck H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii."
RL Science 273:1058-1073(1996).
CC - SIMILARITY: TO PROKARYOTIC TYPE I/III TOPOISOMERASE FAMILY.
DR EMBL: U67605; AAA9673.1;
DR HSSP: P06612; IECL.
DR TIGR: M01652;

Query Match 3.48; Score 98; DB 1; Length 761;
Best Local Similarity 21.58; Pred. No. 13;
Matches 53; Conservative 40; Mismatches 76; Indels 78; Gaps 13.
OY 253 EALTYMTNRREYSLDCK--ENSLFRKQVGVEM-KDKRH-DVTNCRKREI----- 303
DB 288 EAVSYRKIPKETOELAKLXENALISYPRISQKLPKDKKYLEDLINIRNHPYGVKNA 347
OY 304 EETQYQMKRPVPGGYTLQCK-----WITFC 330
DB 348 ERLKENLKP-----VEGKKEDPAHPAHIVDIPKELSEKEKIYDLARRITLAAP 400
OY 331 NOVQDITKINGCKKLYLIGDSTLRQ-W--IYEPKVKVTKLFFDLHETGIFKKHLL 387
DB 401 DMNREYLVNKKIDIGKEFKLSGRVYEGMHEIYFVK-----FEIILPPLKKNKI 453
OY 388 LDMERHTQIOMKHSFPVTFQLSLIDHDYIPREIDRLS-GDKNTAVITFCQHPRPP 446
DB 454 IKVEKITLT-RKETQPKRYTVASII-----KELEKRGLGKMTREI----- 495
OY 447 IDIFIR 453
DB 496 IDKLIR 502

RESULT 14
P78875 PRELIMINARY: PRT: 817 AA.
AC P78875:
DT 01-MAY-1997 (TREMblrel. 03, Created)
DT 01-MAY-1999 (TREMblrel. 10, Last sequence update)
DT 01-MAR-2001 (TREMblrel. 16, Last annotation update)
DE TREHALOSE-PHOSPHATASE (EC 3.1.3.12) (TREHALOSE 6-PHOSPHATE
DE PHOSPHATASE) (TPP).
GN TPPI OR SPAC19612.15C.
OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota: Fungi: Ascomycota: Schizosaccharomycetes:
DB Schizosaccharomycetales: Schizosaccharomycetaceae:
OY NCBL_TaxID=4896:
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX STRAIN=972;
KA Oliver K., Harris D., Wood V., Bartell B.G., Rajendram M.A.;
RT "Isolation and characterization of trehalose-6P phosphatase disrupted
RT mutants from Schizosaccharomyces pombe."
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX STRAIN=972;
KA [3]
RP SEQUENCE OF 505-817 FROM N.A.
RC STRAIN=PR745;
RX MEDLINE=98162722; PubMed=9501991;
RA Yoshioka S., Kato K., Nakai K., Okayama H., Nojima H.;

```

*Identification of open reading frames in Schizosaccharomyces pombe
RT CDNA9.1;
RL DNA Res. 4:363-369(1997).
CC -1- CATALYTIC ACTIVITY: TREHALOSE 6-PHOSPHATE + H(2)O -> TREHALOSE +
CC ORTHOPHOSPHATE.
CC -1- SIMILARITY: CONTAINS A DOMAIN FOUND IN ALL THE PROTEINS OF THE
CC TPS COMPLEX.
DR EMBL; AJ242743; CAB45142.1; -
DR EMBL; Z97209; CAB10126.1; -
DR EMBL; D89225; BAA13886.1; -
DR InterPro; IPR001830; -
DR InterPro; IPR003337; -
DR Pfam; PF00982; Glyco_transf_20; 1.
DR Pfam; PF03356; Trehalose_phrase; 1.
DR Hydrolase.
KM DOMAIN 1 547 TPS COMPLEX DOMAIN.
FT SEQUENCE 817 AA; 93877 MW; 23CFD2533505CBA CRC64;
SQ
Query Match 3.4%; Score 98; DB 3; Length 817;
Best Local Similarity 20.1%; Pred. No. 15; Mismatches 114; Indels 102; Gaps 15;
Matches 66; Conservative 47;
DB 287 KDRKHIDVT--NONKREKIEETCOVGMKPPVGGYTL-----OGKWITTFQNOVO--- 334
113 QDVAFPEETYSOMNGSERSTERYVWMLPGPEKSEITINETSQSGRMALAEVNIPLI 172
OY 335 -----LDTIKINGLCKGL--TYLLGDSLTRO--WYYPKRVYTLK 372
DB 173 HYKTPSSEYDENEDQWMDYVKNNAFPDKICEIKKGDFIYODYSLEVPOLIRNK- 231
OY 373 FFDLHETGIFKHHLLDAE-----KHTQIOMKKHSIPFYFDLYSL----- 413
DB 232 -IDDAVIGFYHHHPSPSELRACPRRAILRSYLGADFIQFEDYSARHFISSCSRYLD 290
OY 414 -IDHDYIPREIDRLSGDKNTAIVITFGHFRRPFI--FIRRA--IGVOKAIEPLFLR 467
DB 291 LEIGHDMV-----NLNGKNTVRAITVG-----IDVPRIHSSGNSVSKLEELNKR 338
OY 468 SPATKVIKNTENIRFEMHETERFGDPFGYI-----HYLINKDI 505
DB 339 YENMKVILGRRLDELGYVOKLRSFORLRTPEMKKKVYLQITISSARKHKLSSI 398
OY 506 FKDLANGIIDAMDNTIAYGTDTHPPDHV 534
DB 399 KK-----LVQA--INQEGTDDYTPVHHV 420
RESULT 15
O9M8K3
ID O9M8K3 PRELIMINARY; PRT: 594 AA.
AC O9M8K3;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE F28L1.9 PROTEIN.
GN F28L1.9.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_Taxid=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV, COLUMBIA;
RA Ilin X., Kaul S., Town C.D., Bentle M., Greasy T.H., Haas B., Wu D.,
RA Rowan C.M., Koo H., Fujii C.Y., Utterback T.R., Barnstead M.E.,
RA Bowman C.L., White O., Nierman W.C., Fraser C.M.;
RT "Arabidopsis thaliana chromosome III BAC F28L1 genomic sequence."
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC018907; AAF30308.1; -
DR SEQUENCE 594 AA; 67574 MW; FDA7ECA139C5CC30 CRC64;

Query Match 3.4%; Score 97.5; DB 10; Length 594;
Best Local Similarity 18.7%; Pred. No. 10;
Matches 128; Conservative 81; Mismatches 255; Indels 207; Gaps 14;
DB 11 LLLISF-SVYTWMIPIIS-QNFKLM--SALNLSIVYNNNSAKSLIFKSLIFKRL 44
DB 31 LTAIVCFMLVWMSIDCSIOSFQVQPRFNAVSVRIS-----PSPSPKSTKINLSKRFH 64
OY 67 TETELRIKEIIEKLDQIIPRPPT--HVNTTSAT-----HSFAI-----LN 14
DB 87 R-----QNLTLMARPNLVKPKNTLITSRVQFVET--AGSKNFTANLAKQWIA 146
OY 108 PROTYCR-----GDQDLLEVRDHLQCKRVQGFPAHSS 140
DB 137 PGAPCKREAKVEISVPGVDISDELAGEIEHFKFQAIDESSKNGVCIQGFEDIS- 195
OY 146 PALTAGASGK---VMDENNGTYLVSTFLWE--GVQSLT--LLIHRSEGAALBARNO 148
DB 196 -----GENMKSRPYKDFGNGTYSFLOVHFPAQDFNLIVILFHHYQGL--FSTSR 244
OY 199 GYKLIIFKGRKFPNGTSHVFTECGILNSNAELCEYLDQGEAFYCMKQHPHCEALTYM 258
DB 249 GFDK-----KLRN-----VALPFV 42
OY 259 TTRNREVSYLTDKENSLEHRSKYGVEMMKDRKHIDVTNCKKREKIEETQVGS- 310
DB 263 KTDVILPELRASCKSPDENRDAWSGHTRLGN-----DEQSLSDNGRFL 409
OY 311 -----MKRPVPG--GYTLOGKI--TFQNOVODLTIKINGLCKGLIYLDS- 355
DB 310 AADFPCKRPMCDGAVGALIESNGVYSSHCSEKFLSARAKWDCKGKWFPMQSDNHVDSI 369
OY 356 -TIRQWYIFFRKVKVTKLFPDL-----HETGIFKHHLLDAERHQTQOMKHS 402
DB 370 RNLNLEVLGHPEIPAVPRRDMKFSMPKPSYVIRISIFNGH-----NMETK 417
OY 403 YPEVFO--LYSLIDHYIPREIDRLSGDKNT--AIVITRCH-----TPP- 445
DB 418 ---NYGLDSLMDRDF--RELLAKYFNEIINRPDVAIVNSGLHSGIHMTSLAATFKA 471
OY 446 -----PIDIFIRAIQVOKAIEPLFLSPAAKVYI1KTENIPEMH 486
DB 472 ETAAAFWEVEVDGKSKSLQPEPEVIFKNTIAT--GGYAFMLAFNPS-----KMEAFN 525
OY 486 ETERFGDFHGYIHYLINKDIFKDLNVGIIIDAMDNTIAYG-----TDTH- 541
DB 526 EK-----MRD--AGLVTSVQDNFQDTYMWIYIDNR--NIGVHYGRVIAKMKH 544
OY 532 DHVIGNO-----INMELNYIC 547
DB 570 DGEIGHQYFVLMVHLVNLALC 592

Search completed: October 12, 2001, 16:05:22
Job time: 165 sec

Fri Oct 12 16:09:19 2001

us-09-729-454-2.rspt

Page 9

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 12, 2001, 16:02:37 ; Search time 29.42 seconds
(without alignments) 1229.875 Million cell updates/sec

Title: US-09-729-454-1

Percent score: 2485
Sequence: 1 MKISMINYKSLALFLIHS.....VHPQHVQNOINLLNYIC 475

Scoring table: RUSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Labelase:
1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1606	64.6	540	2 B45665	adult-specific 61.
2	127.5	5.1	792	2 T00082	hypothetical prote
3	119.5	4.8	3603	1 D69681	peptide synthetase
4	114.5	4.6	807	2 A86740	teichoic acid bios
5	109	4.4	688	2 T33709	hypothetical prote
6	104.5	4.2	4563	1 LPHUB	apolipoprotein B-1
7	104	4.2	694	1 VGBEH6	glycoprotein H pre
8	103	4.1	793	2 S73775	leucine-tRNA 11ga
9	101.5	4.1	1902	1 B44858	lactococin (EC 3.4
10	101.5	4.1	3649	1 S18268	delta-(L-alpha-am
11	101	4.1	988	2 T52134	zwille protein 11m
12	99.5	4.0	1507	2 T42631	breast cancer tumo
13	98.5	4.0	795	2 T07709	hypothetical prote
14	98.5	4.0	1004	2 A71617	SEPA antigen/papal
15	98	3.9	470	2 S54069	hypothetical prote
16	98	3.9	657	2 D81308	probable integral
17	97.5	3.9	634	2 T08810	probable ABC-type
18	97	3.9	538	2 S52472	cell fusion protei
19	97	3.9	905	2 S29329	hypothetical prote
20	97	3.9	1366	2 S57664	IgA-specific metal
21	96.5	3.9	1802	2 T71616	hypothetical prote
22	96.5	3.9	774	1 J00550	1,4-alpha-glucan b
23	96.5	3.9	1245	2 S51255	probable membrane
24	96	3.9	770	2 S76095	hypothetical prote
25	95.5	3.8	1102	2 T23768	hypothetical prote
26	95.5	3.8	1685	2 T02750	probable DNA-direc
27	95.5	3.8	1685	2 T02750	acetyl-CoA carboxy
28	95	3.8	610	1 JC4383	3'-phosphodenosin
29	95	3.8	819	1 T83350	outer membrane ush

30	95	3.8	1021	2 T08601	hypothetical prote
31	95	3.8	1328	1 S04273	retrovirus-related
32	94.5	3.8	740	2 T10619	hypothetical prote
33	94.5	3.8	803	2 T00092	hypothetical prote
34	94	3.8	971	2 A70179	exodeoxyribonucle
35	93.5	3.8	650	2 T41681	probable serine/th
36	93.5	3.8	1209	2 T31657	reverse transcript
37	93.5	3.8	1311	2 T33757	hypothetical prote
38	93.5	3.8	1616	2 S62504	conserved hypothet
39	93.5	3.8	1958	2 T39808	hypothetical prote
40	92.5	3.7	2894	2 C64474	hypothetical prote
41	92.5	3.7	368	2 S46435	porin precursor -
42	92.5	3.7	584	2 H71513	probable sedna exo
43	92.5	3.7	1462	2 T50338	probable guanin-n
44	92.5	3.7	1902	2 B45764	lactococin (EC 3.4
45	92	3.7	462	2 T19231	probable cysteine

ALIGNMENTS

RESULT 1
B45665
adult-specific 61.9K brush border protein precursor - rabbit
C:Species: Oryctolagus cuniculus (domestic rabbit)
C:Date: 03-May-1994 *sequence_revision 03-May-1994 *text_change 04-Mar-2000
C:Accession: B45665
R:Bohl, W.; Schmid-Chanda, T.; Semenza, G.; Mantel, N.
J. Biol. Chem. 268, 12901-12911, 1993
A:Title: Messenger RNAs expressed in intestine of adult but not baby rabbits. Isolati
A:Reference number: M45665; M01D:93286138
A:Accession: B45665
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-540 <BOL>
A:Cross-references: GB:212840; NID:q1761; PIDN:CAA78302.1; PID:q1762
C:Superfamily: rabbit adult-specific 61.9K brush border protein
C:Keywords: intestine

Query Match	64.6%	Score 1606;	DB 2;	Length 540;
Best Local Similarity	57.3%	Pred. No. 7.4e-115;		
Matches 309;	Conservative 67;	Mismatches 91;	Indels 72;	Gaps 2;
QY	8	YKSLALFLIASWIIIFVFNQSKVMSALNLSLHYMNSTKSLFPTPLISKPLTE	67	
DB	3	HKYIKLICLLAICVLCISNSTKIKGALKIPNSHYNTSMSSIFKMSVFKSLTE	62	
QY	68	TELRIKEIIEKLDQOIPRPPTVNTTTSATHTATILNPDYTCRGDHLILEVBDHL	127	
DB	63	TELVEIELEKLDRIIPRPPTVNTTTSATHTATILNPDYTCRGDHLILEVBDYL	122	
QY	128	GRKQGGDFLFRKSSPALMAGASGVDFNNGYIVSFTLFWEGVSLSLIIHPSEG	187	
DB	123	GHQKQGGDFLFRKSSPALMAGASGVDFNNGYIVSFTLFWEGVSLSLIIHPSEG	182	
QY	186	VSALMSRNQGDRIYFTGQFVNGTSQVSEGLITNAELCOYTDNRDQEGFYCYRPQ	247	
DB	183	ASALMSRNQGDRIYFTGQFVNGTSQVSEGLITNAELCOYTDNRDQEGFYCYRPQ	242	
QY	248	HMPCALITMYSKKNKVSYSKQESLFERSNVGVEIMKFNITISVSKCN-----	297	
DB	243	HMPCALITMYSKKNKVSYSKQESLFERSNVGVEIMKFNITISVSKCN-----	297	
QY	298	TLKSYDLHESGKLOHOLAVDLDRNINIMQKTYCPPLIGSMYSVKEMKYLTRAIDRTG	356	
DB	362	KTLEFDLHETGNFKHLLDPAKRNQIQMKKSHSPVYQLFVSDIHGYIPQEDRLIG	421	

```

QY 357 EKNTIVIVISLGGHRRPPIDVFIRALNWKALQHLRLSPDVIYITKTEINREYNDAE 416
      :|||||: |||||||: |||||||: |||||||: |||||||: |||||||: |||||||: ||
Db 422 DKDTIVITEGQHRPPIDIFIRALISVQALERFLRSPATVIVKTEINEMHIAE 481
QY 417 RFSDFHGIVQLIKIDFQDLSVSLIDAMDITAYCTNNVHPPOHVGNOINILNYIC 475
      |||||||: |||||||: |||||||: |||||||: |||||||: |||||||: |||||||: ||
Db 482 RFGDFHGIVQLIKIDFQDLSVSLIDAMDITAYCTNNVHPPOHVGNOINILNYIC 540

RESULT 2
T00082
hypothetical protein KIA0517 - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 22-Jan-1999 #sequence_revision 22-Jan-1999 #text_change 01-Dec-2000
C:Accession: T00082; 114767
R:Nagase, T.; Ishikawa, K.; Miyajima, N.; Tanaka, A.; Kotani, H.; Nomura, N.; Ohara, O.
DNA Res. 5, 31-39, 1998
A:Title: Prediction of the coding sequences of unidentified human genes. IX. The complete
A:Reference number: Z14086; MUID:98290345
A:Accession: T00082
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-792 <NAG>
A:Cross-references: EMBL:AB011089; NID:93043557; PIDN:BA025443.1; PID:93043558
A:Experimental source: brain; clone HG0752
R:Boecker, H.; Boecker, M.; Brandt, P.; Wewes, H.W.; Gassenhuber, J.; Wiemann, S.
Submitted to the protein Sequence Database, August 1999
A:Reference number: Z18182
A:Accession: 114767
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 76,564-792 <BLD>
A:Cross-references: EMBL:AL110234
A:Experimental source: fetal kidney; clone DKFZ566D1824
C:Genetics:
A:Note: KIA0517: DKFZ566D1824.1
A:Superfamily: RING finger homology
F:67-117/Domain: RING finger homology <RNF>

Query Match 5.18; Score 127.5; DB 2; Length 792;
Best Local Similarity 22.18; Pred. No. 0.11;
Matches 89; Conservative 43; Mismatches 133; Indels 137; Gaps 18;
QY 66 TETELRI--KEIIEKL-----DQGIIPRP-----FTHVNTTSTA 97
      |||||: |||||: |||||: |||||: |||||: |||||: |||||: ||
Db 314 TETEVLLVKKQKSEKLNELADQDFPLHPRENDQDFIVETGLKSIHMLGITLTAIVA 373
QY 98 THSTATILNPRDYCRGDQLHILEVYRDLGRKQYGGDFLRARMSPALMAGAGKYTD 157
      :|||: |||: |||: |||: |||: |||: |||: |||
Db 374 SETVATGEGLRQTI--GQPNVVTITTKDQDELCKTGNAVITLDELSTPD--GSVADGELTD 431
QY 158 FNNGTIVYFTLFWEGOVSLSL-----IHP--SEGVSALMSARNQ 197
      |||||: |||||: |||||: |||||: |||||: |||||: |||||: ||
Db 432 NKNGTVEFLTVQKRGDFTLSRLYDQHRGSPFLKVIKRSADVSPTEGCVRRKVSFGS 491
QY 198 GY-----DRYIF-----TGQVNTSQVHSEGS--LIL 223
      |||: |||: |||: |||: |||: |||: |||: |||
Db 492 GHVQKAVKRPASMYSTGKRKNPDEDLIFVYKCGKNGGEFTNIGVASTNKILIIA 551
QY 224 NTNAELCOYLDRND--EGFYCVRRQ-----IHP--CALTHMS-----KKKYSYSKO 270
      :||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||
Db 552 DSNNOCVQIFSDGQFKSRFGIRKSRPGQLQRPQVAVHSGDIIADYDNKNWVSIFSSD 611
QY 271 EKSLFERSNVGEIMEKFNITSVSKCNITLKYDLHESCKLQHLAVDLDRN--INIQMR 348
      :||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||
Db 612 GKF---KTMIG-----SGKLMGPKGVSVDRNGHIIYDVK 643
QY 329 YCYPLIGMTYSVKEMEYLTRAIDRTGKKNVIVISLGGHF 370
      |||: |||: |||: |||: |||: |||: |||: |||
Db 644 ACCVFIFQPRNGKI-----VTRFGSRNGDROFA-----GPHF 675

```

```

RESULT 3
D69681
peptide synthetase ppsd - Bacillus subtilis
C:Species: Bacillus subtilis
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 03-Nov-2000
C:Accession: D69681
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Averydo, V.;
C.: Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capriano, V.; Carter, N.P.;
A.: Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Folia
Nature 390, 249-256, 1997
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, S.; Funa, S.; Galizzi, A.;
leech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.;
Koester, P.; Konigstein, G.; Kroh, S.; Kumano, M.; Kunita, K.; Lapius, A.;
A:Authors: Labbe, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.;
Y., M.; Ogawa, K.; Ogihara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Per
Rieger, M.; Rivolta, C.; Roche, B.; Roche, B.; Rose, M.; Sadleir, V.; Sato, T.; Se
A:Authors: Schleich, S.; Schrieter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpiltra, P.; Tognoni, A.; Tosato, V.;
T.: Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasunori, K.; Yata, K.; Yosh
A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Zumbstein, A.
A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis
A:Reference number: A69580; MUID:98044033
A:Accession: D69681
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-3603 <KUN>
A:Cross-references: GB:299113; GB:AL009126; NID:92634090; PIDN:AB1714.1; PID:92
A:Experimental source: strain 168
C:Genetics:
A:Gene: ppsd
A:Superfamily: surfactin synthetase; acetate--CoA lyase homology; acyl carrier protein
C:Keywords: carrier protein; phosphopantetheine; phosphoprotein
F:509-992/Domain: acetate--CoA lyase homology <ACLI>
F:969-1037/Domain: acyl carrier protein homology <ACLP>
F:1540-1983/Domain: acetate--CoA lyase homology <ACLD>
F:2000-2068/Domain: acyl carrier protein homology <ACPF>
F:2579-1019/Domain: acetate--CoA lyase homology <ACU3>
F:3037-1104/Domain: acyl carrier protein homology <ACU3>
F:1001-2032,3069/Binding site: phosphopantetheine (Ser) (covalent) #status predicted

Query Match 4.88; Score 119.5; DB 1; Length 3603;
Best Local Similarity 20.78; Pred. No. 3.9;
Matches 98; Conservative 73; Mismatches 149; Indels 153; Gaps 2;
QY 23 IFVFNSTKY-----MSALNISLSHYNNKNTSLPKPKPLSLKLTLEFLRIKE 74
      |||: |||: |||: |||: |||: |||: |||: |||
Db 1426 LFCQFESTALEKETTIOKMSLETTLVH-----TRASPELELDNIPLTKREK--D 1477
QY 75 IIR-----LDQGIIPRP-----FTHVNTTSTAIDTAII 157
      |||: |||: |||: |||: |||: |||: |||: |||
Db 1478 FIESCHLPFEETGYSNQTLYHLEQDAEKTPOAAVIFEDGVATYKELNEUARISEML 1647
QY 106 NPRDYCRGDQLHILEVYRDLGRKQYGGDFLRARMSPALMAGAGKYTD 157
      |||: |||: |||: |||: |||: |||: |||: |||
Db 1538 G-----RGVKETVAI---IKR-----SPEMLDGIYGLK--AGGATIP 1573
QY 166 SPFLFWEGOVSLSLIHPSEGVSALMSARNGDQRYVITFGVNTSQVHSEGS--LIL 223
      |||: |||: |||: |||: |||: |||: |||: |||
Db 1574 IDPDYPERI--SELLDSGTNILLQSA--GLHVEFGTEIV-----YLNLT 1617
QY 226 NAELOCYLDNRDQEGFYCVRRQHPICALTHYSKNNKRVYLSKUEKSLFERSNV--VEIM 248
      |||: |||: |||: |||: |||: |||: |||: |||
Db 1618 NSGLAHLRSLNPVD---VLPQSLAVVIVTSSTGMR-----GVEIE 1656
QY 286 EKFNITSVSKCNITLKS--VDHESGKIQHLAVDLDNRNINQWOKKYCYPLIGMTYSV--- 441
      |||: |||: |||: |||: |||: |||: |||: |||
Db 1657 HR---SAVNFSLQSRVQKHSQDMIMHTSIFSDASI--WELFWKVPYAAASVYIILNQ 1710
QY 342 --KEMEYLTRAIDRTGKKNVIVISLGGHRRPPIDVFIRALNWKALQHLRLSPD 399
      |||: |||: |||: |||: |||: |||: |||: |||
Db 1711 GEKEPEVIAKAIIE---QKITAM-----HVPVSMU-----HAFLEQIKVRS-- 1748

```


OY 400 MYIITENIREMYNDAREFSDFGYIOYLIIKIDIFODLSVSIIDAMDTIATG 452
 DB 1749 -VPIKIMKLKRVFSGGEOLG-----THLYSR--FTLELPPN-----SITSYG 1788

RESULT 4

teichoic acid biosynthesis protein [imported] - Lactococcus lactis subsp. lactis (strain G-32)
 C:Species: Lactococcus lactis subsp. lactis
 C:Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 23-Mar-2001
 C:Accession: AB6740
 R:Bohlin, A.; Muncher, P.; Manger, S.; Jallion, O.; Malarme, K.; Weissenbach, J.; Ehrlich, S. In press, 2001
 A:Title: The complete genome sequence of the lactic acid bacterium.
 A:Reference number: AB6625
 A:Accession: AB6740
 A:Status: Preliminary
 A:Molecule type: DNA
 A:Residues: 1-807 <STO>
 A:Cross-references: GB:AF005176; MID:q12723855; PIDN:AAK05019.1; GSPDB:GN00146
 A:Experimental source: strain IL1403
 C:Genetics:

Query Match 4.6%, Score 114.5; DB 2; Length 807;
 Best Local Similarity 20.0%, Pred. No. 1.1;
 Matches 90; Conservative 76; Mismatches 175; Indels 109; Gaps 18;

OY 29 NSTVMSALMLSLIHWNNSTKSLFPTLISLKLPIETELRIKEITEKLDQIIPR 85
 DB 160 HTIKFAGSGCYKRGIVYGEIIESGYRIDLTKETSSDANKLKEGIELDSNPVYVY 219
 OY 86 KPTHTNTTISATHTATILNPRDTCRGDQ---HILEVRDLGR---KQYGGFL 138
 DB 220 MPTWGNQDTONPDSIAQVI-AELKYLKKEFLGKYNIIILKHPYLKAKANKKELSGVLI 278
 OY 139 RARMSPALMAGASGVTDENNGYLVSPFLMEQGVSLIILHPSEGSALMSRRNG 198
 DB 279 DQAMDAVLVAATDILITDFS-----SVFPDYLVTDKPIFY-----AMDDIYS 323
 OY 199 YDVIPTGQFVNGTQVHSECGLLNTMELCOYLDNDGCFYCVAPDHPCALTHMY 258
 DB 424 EDR---GMFLD-MELPAP---ILKTVIEIADYLSDDO-----L 356
 OY 259 SKNKVSVLSKQEKSL-FERSNVGVEIM-----EKFTISVSKNTLKVSDCHESGKLQ 311
 DB 357 SDYLGKYLAKKEXYVYDQGNSEIRIVQYIFRKKESQLVYKKIDSEKILLFTPGMID 416
 OY 312 HOLAVDLDRNIN--IQOKCYCPIL-----GSMITSVKEMEYLTRAIDRTG---GEKNT 360
 DB 417 NNGITOSFINTLNALDYOKYVIVFTNTPKSHFPNNYQKLKKNIRLIFRTGSPNFESEKQ 476
 OY 361 VI--VISTAGHERPPPIVFIIRA-----LNVHKAIOHLRLSPDTMVIITENIREMY 412
 DB 427 MIMHKKIKSKSHITSLPEVAFFKREHRLFGSLSPKAI----- 513
 OY 413 NDAERESDFHGYIOYLIIKIDIFODLSVSI 442
 DB 514 -----DFSGISFYWSKTFVAFSDSRVKM 536

RESULT 5

hypothetical protein F58E2.3 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 08-Sep-2000
 C:Accession: T33709
 R:Goebl, D.; Telehaunty, A.
 Submitted to the EMBL Data Library, October 1998
 A:Description: The sequence of C. elegans cosmid F58E2.
 A:Reference number: Z21490
 A:Accession: T33709

A:Status: Preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-688 <GCE>
 A:Cross-references: EMBL:AF100659; PIDN:AMC68966.1; GSPDB:GN00022; CESP:F58E2.3
 A:Experimental source: strain Bristol NZ; clone F58E2
 C:Genetics:
 A:Gene: CESP:F58E2.3
 A:Map position: 4
 A:Introns: 227/3; 306/3; 341/2; 602/3
 C:Superfamily: Caenorhabditis elegans hypothetical protein F58E2.3

Query Match 4.4%, Score 109; DB 2; Length 688;
 Best Local Similarity 19.3%, Pred. No. 2.3;
 Matches 112; Conservative 76; Mismatches 200; Indels 192; Gaps 27;

OY 1 MKISIMVKSLLALLFIASMTITTVQNSTKYMSNLISLHW-----NSTKSLF 54
 DB 93 LEISQDIIEEL-----DKMMSASSQSIQDFWDFIPLNNHRTGGK 133
 OY 55 KPTPLISL-----KPLTETELRIKEITEKLDQIIPR 86
 DB 134 PK--LISLFGTQGLSIFRHSNHLIINLFRKEDTEIEHFQETVKM--KQKHA 189
 OY 87 PETHTNTTISATHTATILNPRDTCRGDQ---HILEVRDLGRKQYGGFLAARMS 144
 DB 190 KYHLDEECTSISNVLHFGANFVYDISDQLMKELSTHSAITLYRNIISF----- 242
 OY 145 PALMAGASGVTD-----FNNGYLVSTLFW--EGVYSLIILHPBGV---SALMS 193
 DB 243 ----NNMADSDITDESITGLFSGPKRHSLEFMSWESIDGKIECFSEFPKKFSIYCPKACE 298
 OY 194 ARNOGYDRVIFTGQFVNGTQVHSECG---GILMT-----NAELCOYLDNRQDG 240
 DB 299 LRN-----VQSKELCETSICLSCTSELSIKYINGLKMLNLSHGRIDCL--RTSRG 350
 OY 241 FYCYRPHMPCALTHMYSKKKVSVLSKQESLERSNVGVEIMEKFTISVSC----- 296
 DB 351 V-----QKCVDTI-----KPAETIAEINLNVENVSVI---FMAKIKKATYEP 394
 OY 297 -----NTKSYDLHESGKLOHOLAVDLDRNINIQM-----KYCPLIGSMT 338
 DB 395 IDLTTSYQDVHVKVIOKVEKHRNGLVERELC-----SVNNMQGFSLKFADLKLMI 449
 OY 339 YSVKEME-----YLTRAIDRTGGEKTYIVY-----SLGQHERPPFIQYFI 379
 DB 450 AKSEIQCYNLDPSILKANSRPGSKNKHVLLRPLAVQOLDVISTTQDQNLQIVPKLS 509
 OY 380 RRALNVKRAIOHLRLSPDTMVIITENIREMYNDAREFSDFGYIOYLIIKIDFODLSV 439
 DB 510 IOVNLQTLIOLSLRPS--LKTIEIKKLND-----IISKSLKLEA 553
 OY 440 SITDAW---DITIAVGTNNVHPQ-----HVVQMIIL 470
 DB 554 AQDDQMKNMEDILTEVL---APQIOYFLHVNTEINIM 590

RESULT 6

LPHOB
 apolipoprotein B-100 precursor - human
 N:Contains: apolipoprotein B-26; apolipoprotein B-48; apolipoprotein B-74
 C:Species: Homo sapiens (man)
 C:Date: 28-Dec-1987 #sequence_revision 28-Dec-1987 #text_change 21-Jul-2000
 C:Accession: A27850; A25679; A25263; A25266; A24320; A24684; A23817; A25774
 A452; I61909; I59510; I39474; I39469; I84624; I37179; P50058
 R:Lundkvist, E.H.; Blackhart, B.D.; Pleroff, V.R.; Catali, L.; Fortier, C.; Knott, T.
 DNA 6, 363-372, 1987
 A:Title: DNA sequence of the human apolipoprotein B gene.
 A:Reference number: A27850; MID:88003974
 A:Accession: A27850
 A:Molecule type: DNA
 A:Residues: 1-617 'A', 619-1929, 'F', 1931-3318, 'D', 3320-3426, 'T', 3428-3431, 'Q', 3433-3
 A:Cross-references: GB:M4162

R:Cladaras, C.; Hadzopoulou-Cladaras, M.; Nolte, R.T.; Atkinson, D.; Zannis, V.I.
EMBO J. 5, 3495-3507, 1986
A:Title: The complete sequence and structural analysis of human apolipoprotein B-100: re
A:Reference number: A91058; MUID:87161758
A:Accession: A25679
A:Molecule type: mRNA
A:Residues: 1-11, 15-2539, 'S', 2541-3823, 'R', 3825-4563 <CLA>
A>Note: 1109-Asp was also found
R:Knott, T.J.; Wallis, S.C.; Powell, L.M.; Pease, R.J.; Lusis, A.J.; Blackhart, B.; McCa
Nucleic Acids Res. 14, 7501-7503, 1986
A:Title: Complete cDNA and derived protein sequence of human apolipoprotein B-100.
A:Reference number: A93639; MUID:87016385
A:Accession: A25263
A:Molecule type: mRNA
A:Residues: 1-272, 'N', 274-617, 'A', 619-1217, 'E', 1219-2091, 'V', 2093-2364, 'T', 2366-2679, 'Q'
A:Cross-references: GB:X04506; NID:934330; PIDN:CAA28191.1; PID:934331
R:Law, S.W.; Grant, S.M.; Hlouchil, K.; Hospattankar, A.; Lackner, K.; Lee, N.; Brewer JH
Proc. Natl. Acad. Sci. U.S.A. 83, 8142-8146, 1986
A:Title: Human liver apolipoprotein B-100 cDNA: complete nucleic acid and derived amino
A:Reference number: A94134; MUID:87041416
A:Accession: A25267
A:Molecule type: mRNA
A:Residues: 1-617, 'A', 619-703, 'P', 705-792, 'R', 794-1270, 'S', 1272-1866, 'G', 1868-2036, 'N', 2
4189-4220, 'W', 4222-4563 <LAW>
A>Note: The codons given for residues 704, 793, 1271, 2037, 2933, 3286, 3782, 4188, and
R:Chen, S.H.; Yang, C.Y.; Chen, P.F.; Setzer, D.; Tanimura, M.; Li, W.H.; Gotto Jr., A.H
J. Biol. Chem. 261, 12918-12921, 1986
A:Title: The complete cDNA and amino acid sequence of human apolipoprotein B-100.
A:Reference number: A92556; MUID:87008488
A:Accession: A25266
A:Molecule type: mRNA
A:Residues: 1-97, 'T', 99-328, 'V', 330-644, 'T', 646-918, 'P', 920-3318, 'D', 3320-3426, 'T', 3428-
9-4132, 'G', 4134-4180, 'E', 4182-4563 <CHS>
A:Cross-references: GB:J02610; NID:9178803; PIDN:AAA5549.1; PID:9178804
A>Note: A total of 2366 residues were confirmed by direct sequencing of tryptic peptides
R:Protter, A.A.; Hardman, D.A.; Seto, K.Y.; Schilling, J.W.; Yamanka, M.; Holt, T.J.; H
Proc. Natl. Acad. Sci. U.S.A. 83, 5678-5682, 1986
A:Title: Analysis of cDNA clones encoding the entire B-26 region of human apolipoprotein
A:Reference number: A24320; MUID:86287319
A:Accession: A24320
A:Molecule type: mRNA
A:Residues: 1-97, 'T', 99-617, 'A', 619-941, 'YYINSLPPKP', 951-1138, 'PTGRLPNCFNGLCYSLSLHSPFQ
A:Cross-references: GB:M12480; NID:9178791; PIDN:AAA51791.1; PID:9178792
R:Protter, A.A.; Hardman, D.A.; Schilling, J.W.; Miller, J.; Appleby, V.; Chen, G.C.; K
Proc. Natl. Acad. Sci. U.S.A. 83, 1467-1471, 1986
A:Title: Isolation of a cDNA clone encoding the amino-terminal region of human apolipop
A:Reference number: A94088; MUID:86149925
A:Accession: A23817
A:Molecule type: mRNA
A:Residues: 485-617, 'A', 619-1044 <LA2>
A:Cross-references: GB:M12480; NID:9178791; PIDN:AAA51791.1; PID:9178798
R:Deeb, S.S.; Koculsky, A.G.; Albers, J.J.
Proc. Natl. Acad. Sci. U.S.A. 82, 4983-4986, 1985
A:Title: A partial cDNA clone for human apolipoprotein B.
A:Reference number: A25774; MUID:85270450
A:Accession: A25774
A:Molecule type: mRNA
A:Residues: 709-791, 'SSMKASHGCPHSGN', 810-906 <DEE>
A:Cross-references: GB:K03175; NID:9178821; PIDN:AAA51759.1; PID:9178822
R:Carlsson, P.; Darnofre, C.; Olofsson, S.O.; Bjursell, G.
Gene 49, 29-51, 1986
A:Title: Analysis of the human apolipoprotein B gene: complete structure of the B-74 reg
A:Reference number: A91565; MUID:87191999
A:Accession: A25533
A:Molecule type: mRNA

A:Residues: 1282-2721, 2742-3290, 'V', 3292-3336, 'N', 3338-3348, 'P', 3350-3363, 'V', 3365-3375, 'V'
A:Cross-references: GB:M15421; NID:9178817; PIDN:AAA51758.1; PID:9178818
R:Hardman, D.A.; Protter, A.A.; Chen, G.C.; Schilling, J.W.; Seto, K.Y.; Li, W.H.
Biochemistry 26, 5478-5486, 1987
A:Title: Structural comparison of human apolipoproteins B-48 and B-100
A:Reference number: A29671; MUID:86050832
A:Accession: A29671
A:Molecule type: mRNA
A:Residues: 1671-2133, 'PYW', 2327-2352, 'H', 2354-2398, 'HAB',
A:Cross-references: GB:M17367; NID:9178731; PIDN:AAA51742.1; PID:9178732
R:Shoulters, C.C.; Myant, N.B.; Siddall, A.; Rodriguez, J.C.; Cortese, G.; Battist
Atherosclerosis 58, 277-289, 1985
A:Title: Molecular cloning of human LDL apolipoprotein B cDNA. Evidence for two
A:Reference number: A90084; MUID:86130855
A:Accession: A29287
A:Molecule type: mRNA
A:Residues: 3846-4298 <SHO>
R:Pitzner, R.; Wagener, R.; Stoffel, W.
Biol. Chem. Hoppe-Seyler 367, 1077-1083, 1986
A:Title: Isolation, expression and characterization of a human apolipoprotein B
A:Reference number: A25572; MUID:87076044
A:Accession: A25572
A:Molecule type: mRNA
A:Residues: 4219-4337, 'S', 4339-4563 <PEI>
A:Cross-references: GB:M36676
R:Wei, C.F.; Chen, S.H.; Yang, C.Y.; Marcel, Y.L.; Milne, R.W.; Li, W.H.; Sparrow
Proc. Natl. Acad. Sci. U.S.A. 82, 7265-7269, 1985
A:Reference number: A24738; MUID:86042646
A:Accession: A24738
A:Molecule type: mRNA
A:Residues: 'N', 3729-3731, 'T', 3733-3875, 'A', 3877-3948, 'P', 3950-3963, 'V', 3965-3975, 'V'
A:Cross-references: GB:M12413; NID:9178735; PIDN:AAA51742.1; PID:9178736
R:Chen, S.H.; Haddil, G.; Yang, C.Y.; Gu, Z.W.; Lee, B.K.; Weig, S.J.; Schilling, J.
Science 238, 363-366, 1987
A:Title: Apolipoprotein B-48 is the product of a messenger RNA with an unusual 5' end
A:Reference number: A40133; MUID:86018019
A:Accession: B40133
A:Molecule type: mRNA
A:Residues: 2165-2179 <CH1>
A:Cross-references: GB:M18036; NID:9178799; PIDN:AAA51754.1; PID:9178800
A>Note: this mRNA includes the stop codon of the organ-specific mRNA for apoB
A:Accession: A40133
A:Molecule type: protein
A:Residues: 51-75, 101-110, 129-139, 158-174, 197-207, 276-287, 298-304, 304-313, 316-324,
363-486, 1498-1556, 1563-1572, 1601-1610, 1647-1661, 1697-1724, 1770-1781, 1879-1897
A>Note: these fragments were derived from apoB
R:Hardman, D.A.; Protter, A.A.; Schilling, J.W.; Kane, J.P.
Biochem. Biophys. Res. Commun. 149, 1214-1219, 1987
A:Title: Carboxyl terminal analysis of human B-48 protein confirms the novel method
A:Reference number: A28002; MUID:86106542
A:Accession: A28002
A:Molecule type: mRNA
A:Residues: 2129-2179, 2181-2235 <HA2>
A:Cross-references: GB:M18471
A:Experimental source: intestine
A>Note: this mRNA from intestine includes a stop codon created by RNA editing in
Nucleic Acids Res. 13, 6937-6953, 1985
A:Title: Human apolipoprotein B: identification of cDNA clones and characterization
A:Reference number: A24269; MUID:86041886
A:Accession: A24269
A:Molecule type: mRNA
A:Residues: 3056-3159 <MEH>
A:Cross-references: GB:X03045; NID:928783; PIDN:CAA24850.1; PID:929609
R:Hospattankar, A.V.; Hlouchil, K.; Law, S.W.; Meglin, N.; Brewer JH.
Biochem. Biophys. Res. Commun. 148, 279-285, 1987
A:Title: Identification of a novel in-frame translational stop codon in human apoB
A:Reference number: A29659; MUID:8604670
A:Accession: A29659
A:Molecule type: mRNA
A:Residues: 2169-2179 <HOS>
A>Note: the sequence shown represents the carboxyl end of apolipoprotein B-48
A>Note: two RNA species, 14.1kb and 7.5kb in length, were isolated from the human

```

Db      2395 -----RVLDDQLGT-----TFSERI---NDVL-----EHVKNFVINLI 2325
Oy      376 -DVFIRRLNHNKAIQHILLRSPDT-----MVIIKTENIREMYNDACRSDFPHGIQYLI 429
Db      2326 GDEFEVERIKINAFRAKHVELIEREVEDQIOVLMDKLVELANQYLKTKTIQKLSNVLDQYK 2385
Oy      430 IKDFIDOLSVSIIDA 444
        ||| | : | ||
Db      2386 IKDFEKLVEGFIDDA 2400

RESULT 7
VCSEB6
glycoprotein H precursor - human herpesvirus 6 (strains U1102, GS, and AJ)
C:Species: human herpesvirus 6
C:Date: 31-Dec-1992 #sequence-revision 31-Dec-1992 #text-change 24-Sep-1999
C:Accession: A40511; J02382; P00846; D56653; P00847
J:Josephson, S.F.; Ablashi, D.V.; Salahuddin, S.Z.; Jagodzinski, L.L.; Wong-Staal, F., et al.
J:Virology 65, 5597-5604, 1991
A>Title: Identification of the human herpesvirus 6 glycoprotein H and putative largep
A:Reference number: A40511; MUID:91374623
A:Accession: A40511
A:Molecule type: DNA
A:Residues: 1-694 <GOS>
A:Cross-references: GB:S57509; NID:g235428; PIDN:AAB19779.1; PID:g335431
A:Experimental source: strain GS, clone pZYBA3
R:Gompels, U.A.; Carrigan, D.R.; Carss, A.L.; Arno, J.
J:Gen. Virol. 74, 613-622, 1993
A>Title: Two groups of human herpesvirus 6 identified by sequence analyses of laborat
A:Reference number: J02382; MUID:93224882
A:Accession: J02382
A:Molecule type: DNA
A:Residues: 1-694 <GOM>
A:Cross-references: GB:X83413; NID:g853961; PIDN:CAA58382.1; PID:g854027
A:Experimental source: strain AJ
A:Accession: P00846
A:Molecule type: DNA
A:Residues: 1-65 <GOM1>
A:Experimental source: strain GS
R:Gompels, U.A.; Carss, A.L.; Sun, N.; Arrand, J.R.
DNA Seq. 3, 25-39, 1992
A>Title: Infectivity determinants encoded in a conserved gene block of human herpev
A:Reference number: A56653; MUID:93091236
A:Accession: D56653
A:Molecule type: DNA
A:Residues: 1-694 <G02>
A:Cross-references: GB:X83413; NID:g853961; PIDN:CAA58382.1; PID:g854027
A:Experimental source: strain U1102
A>Note: sequence extracted from NCBI backbone (NCBI:N120533, NCBI:P121866)
C:Superfamily: herpesvirus glycoprotein H
C:Keywords: glycoprotein; transmembrane protein
F:1-23/Domain: signal sequence
F:24-694/Product: glycoprotein H #status predicted <SIG>
F:554-570/Domain: transmembrane #status predicted <TM1>
F:668-684/Domain: transmembrane #status predicted <TM2>
F:721,24,51,76,148,206,337,365,384,411,538,573,593,652/Binding site: carbohydrate (As
Query Match 4.2% score 104; DB 1; Length 694;
Best local similarity 19.4%; Pred No. 5,6;
Matches 101; Conservative 77; Mismatches 208; Indels 134; Gaps 24;

Oy      17 IIASTIIFYFONSTKWSALNTSISLHTWNNSTKSLEPTPLISLKPLETELRKEII 76
Db      2 LIRLWPF-VLLTPCYGNRPILNISNSHCRNGFNENPIVRPGIFTFNFYTKNDRIYOVP 59
Oy      77 E-KLDQOIIPRPFTYHWNTTSATSHATILN----PROTCRDQDHILLEVDNHGRKK 131
Db      60 KCLLAGSDITVHLFDAINTTESLTNYEKRVTRPREPPMN----- 97
Oy      132 QYGGPFLRAKVSSPALMGASGVYTPPNNGITVSLFLPEQGYSLSLIHHSQEV--S 189

```

```

QY 190 ALMSARNOGYDRVIFTGQFVNGTSG---VHSEGGILNTNAELCOYLDNRDGEFCYR 245
D 137 RVEVYRQMGYDNV---SCKLPNSKKEIFFPYVRC-----AKITRYGEDIYTHFF--T 185
QY 246 PQHM-----PCAALTHMSKKKKVSYLSKQES--LFRSNVNG-----VEIMEK 287
D 186 PQEMILYTONPAGDLTMMYGNTTSINFKAPYKSSFFKQTLTDLLILEKNDIYQYR 245
QY 288 FNTISVSKCNTLKSVDLHESGRLQHLAVDLDRINIQ-----MOKYCPILGSR--TY 339
D 246 FTSDFATFVDETLDNDVDEVELLKFE-----NNLDIQLLRGGCKKPNVAGIPQMEFLY 298
QY 340 SVKEMEYLTIRAIIDRTGSEKNT---VIYSLGQHFPPFDIFIRALNVH-----KAI 389
D 299 GIVHFSYST-----KNTGPMFVLRLKTHENLISDFSVKCVNVEGTLQYPMKK 349
QY 390 QHLLRSPDTNVIKTENT-----REMYNDAERFSDFGYIQTLYIKIDFQ 435
D 350 EFLKYEPSDYSLTKNKSISVSTLTLYLATAYESNTLSKYKMTDIANTQNYEKHMF- 408
QY 436 DLVSIIIDAMDITIAVGTNNVHPQHYVNGQIILLNTIC 475
D 409 FNTLFSDBRETLFMAELNITPIDERMRHQLIGNDC 448

```

RESULT 8
 S73779
 leucine--trna, ligase (EC 6.1.1.4) leus - Mycoplasma pneumoniae (strain ATCC 29342)
 N:Alternate names: hypothetical protein F11_07f930; leucyl-trna synthetase leus
 C:Species: Mycoplasma pneumoniae
 A:Variety: ATCC 29342
 C:Date: 27-Feb-1997 #sequence-revision 25-Apr-1997 #text-change 07-Dec-1999
 C:Accession: S73779
 R:Himmelreich, R.; Hilbert, H.; Plagens, H.; Pirkil, E.; Li, B.C.; Herrmann, R.
 Nucleic Acids Res. 24, 4420-4449, 1996
 A:Title: Complete sequence analysis of the genome of the bacterium Mycoplasma pneumoniae
 A:Reference number: S73327; MUID:97105885
 A:Accession: S73779
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-793 <HIM>
 A:Cross-references: EMBL:AE000044; GB:000089; NID:g1674130; PIDN:AAB96101.1; PID:g167414
 A>Note: the nucleotide sequence was submitted to the EMBL Data Library, November 1996
 C:Genetics:
 A:Gene: leus
 A:Genetic code: SGC3
 C:Superfamily: leucine--trna ligase
 C:Keywords: aminoacyl-trna synthetase; ligase; protein biosynthesis

Query Match 4.1%; Score 103; DB 2; Length 793;
 Best Local Similarity 19.2%; Pred. No. 8.1;
 Matches 86; Conservative 85; Mismatches 174; Indels 102; Gaps 20;

```

QY 1 MKISMTNKSLLALLFTLASMIIFTVFONSTRKVSALNLSISLHYNNSTSLPPTPLI 60
D 306 LKTKAINPAINTAIPLYVANYVEGTDVAV-----MGVPAHNENNPFAKAKQALPIT 358
QY 61 SLKPLTELELKEIEIKLDQOIPRPPTNNTTSTSTSTITINPDTYCRQDQHLIL 120
D 359 T-----VIDKQERLQHSQTSGLNSOTANTQITMLVERKA-----KKTIV 400
QY 121 LEVRDHLRRKQYGDFLRARSSPALMAGASGVTDENNQTYLVSFTLPEWQVSLSL 180
D 401 YKLRMIFSRQRYWGE-----PPIPLF-----DENNGHLVK-----ELPVLPLRL 441
QY 181 LIHPSEGV--ALMSARNOGYDRVIFTGQ--FVNGTS--QVNSC-----GLINTNAELIC 230
D 442 ANYOPDGTNPPLM--RNOEMAKVQGNQITRETSTPOMWAGSCWYLLGYLMLINNEIF 499
QY 231 QYLDNRDGEFCYRPOHMF-----AALTHMSKKKKVSYLSKQESL 274

```

```

D 500 WEIDREAKDLF---ERYLPVDLYVGAENAVJHLXAMPQHPFYCKGVTTKEFEKL 544
QY 275 FPRSNVGEIMEFENFTIISKCNITKS---VDLHESGKLRLAVDLFNTINQWLRVY 611
D 557 T---NQNVLAPDQKMSKSKGNITINPTPLIDSHGADALRYL-----MIMG 661
QY 332 PLIGSMYSVKEMEYLTIRAIIDRTG--EKNTVISLGHFFPFTVPIRLANVHMF 694
D 601 PISALITMLDDGNGRRLDVRVINEFHKENIIKETVDE--TVVYNLFLKNSFE-HLK 650
QY 390 QHLLRSPDTNVIIT---KTENIREMY 412
D 659 QELNLVYSQMKHIFLNLKTKQLTLAY 685

```

RESULT 9
 B44858
 lactocepin (EC 3.4.21.96) precursor [validated] - Lactobacillus paracasei (strain N:Alternate names: cell-envelope-associated proteinase pitp; serine proteinase" Lf
 C:Species: Lactobacillus paracasei
 C:Date: 10-Sep-1999 #sequence-revision 10-Sep-1999 #text-change 19-May-2000
 C:Accession: B44858; C44858; A44850
 R:Holck, A.; Naes, H.
 J. Gen. Microbiol. 138, 1353-1364, 1992
 A:Title: Cloning, sequencing and expression of the gene encoding the cell-envelope
 A:Reference number: A44858; MUID:92381481
 A:Accession: B44858
 A:Molecule type: DNA
 A:Residues: 1-1902 <HOL1>
 A:Cross-references: GB:M83946; NID:g149580; PIDN:AAA45446.1; PID:g149442
 A>Note: sequence extracted from NCBI backbone (NCBI:112261, NCBI:P112261.4)
 A:Accession: C44858
 A:Molecule type: protein
 A:Residues: X', 189-196 <HOL2>
 R:Naes, H.; Nissen-Meyer, J.
 J. Gen. Microbiol. 138, 313-318, 1992
 A:Title: Purification and N-terminal amino acid sequence determination of the cell-en
 A:Reference number: A44850; MUID:92226694
 A:Accession: A44850
 A:Status: preliminary
 A:Molecule type: protein
 A:Residues: 189-196 <NAE>
 A:Cross-references: PIDN:AAB22052.1; PID:g248666
 A:Experimental source: strain NCDO 151
 A>Note: sequence extracted from NCBI backbone (NCBI:P94706)
 C:Genetics:
 A:Gene: pitp
 C:Superfamily: lactocepin; subtilisin homology
 C:Keywords: hydrolase; serine proteinase; transmembrane protein
 F:1-23/Domain: signal sequence
 F:23-187/Domain: propeptide
 F:188-1902/Product: serine proteinase, cell-envelope-associated
 F:208-634/Domain: subtilisin homology #status atypical <SBI>

Query Match 4.1%; Score 101.5; DB 1; Length 1902;
 Best Local Similarity 21.6%; Pred. No. 37;
 Matches 102; Conservative 47; Mismatches 159; Indels 165; Gaps 24;

```

QY 58 PLISAKPLTELELRK-----ETLEKLDQOIPRPPTNNTTSTSTSTITINPDTYCRQDQHLIL 108
D 714 PAVELKQDFTSDTFTFLPTNTHLTYOMDS-----NIDTNAVYTSATDPN-- 761
QY 109 DTYCRQDQHLILLEVRDHLGRKKQYGDFLRARSSPALMAGASGVTDENNQTYLVSFTLPEWQVSLSL 168
D 762 -----SGVLDKKIIDDAIKAGSD--ITVPAKTAQIEFT 794
QY 169 L-----FMQGVSLSLLIHPSEGVSLMSARNOGYDRVIFTGQFVNG-----TSOVHS 217
D 795 LSKRSPDQQFVEGFLNFKSDG-----SRLLPLP--MGFFGDMWNGKIVDSLNGITYS 847
QY 218 ECG-----LILNTNAELCOYLDN--RDGEFCYRPOHMFCAALTHMSKKKKVSYLS 266

```

```

Db 848 PAGNGTGVPLLTUKNTNG-HQYIGGAVTDAOSKQTVDDQ-----ALAFSSDKMALYNDIS 901
QY 269 KQKSLFERNVGEVIME-----KFNITSKCNLTLSVDLHESGKLQHLQAVLDNRINI 324
Db 902 MUYLLRNISNOVDILDGQGNKVITLLSSSTNQKRYVDAS----- 943
QY 325 QWQKCYC---PL-----IGSMYSVKEMEYLTRAIDRTGKNTYIVI 364
Db 344 ---QKTYVNAFAMQSTYYDQKQCNIKTADDSYTRISGVP-----EGSDKQVFDV 993
QY 365 SLGNHFRPFTIVFIRALNVHKAIOHLLNSPDITWITENIR-EMYNDAERSPFHG 423
Db 374 -----FFKLD-----SKAPYRHWALSA-----KTEGKQYVLYLTAEKKDLSC 1032
QY 424 YIQYLLIKDIEQD--LSVSIIDAMDITAYGTNNVHP-----QHVVGNQIN 468
Db 1033 LIAKRSVKTAINEVTLNLCATFTDAG--TTAAGYTKIETPLSDQAQAGNNDN 1083

RESULT 10
S18268
delta-(D-alpha-aminoadipyl)-L-cysteinyl-D-valine synthetase - Streptomyces lactandurans
C:Species: Streptomyces lactandurans
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 03-Nov-2000
C:Accession: S18268; S15283; B38171
R:Martin, J.F.
submitted to the EMBL Data Library, January 1991
A:Reference number: S18268
A:Accession: S18268
A:Molecule type: DNA
A:Residues: 1-3649 <MAR>
A:Cross-references: EMBL:X57310; NID:q45005; PIDN:CAA0561.1; PID:q45006
K:Ogawa, J.J.R.; Martin, J.F.; Calzada, J.G.; Liras, P.
Mol. Microbiol. 5, 1125-1133, 1991
A:Title: The cephamycin biosynthetic genes pcbB, encoding a large multidomain peptide
genes in Actinobium chrysogenum and Penicillium chrysogenum.
A:Reference number: S15283; MOID:92063608
A:Accession: S15283
A>Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 224-940/119-2010/2373-3307 <COQ>
A:Cross-references: EMBL:X57310
A:Note: the source is designated as Nocardi lactandurans
K:Ogawa, J.J.R.; Liras, P.; Laiz, L.; Martin, J.F.
J. Bacteriol. 173, 6258-6264, 1991
A:Title: A gene encoding lysine 6-aminotransferase, which forms the beta-lactam precursor
A:Reference number: A8171; MOID:92011990
A:Accession: B38171
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-21 <CO2>
A:Cross-references: GB:S57006
C:Genetics:
A:Gene: pcbAH
A:Superfamily: alpha-aminoadipyl-cysteinyl-valine synthetase; acetate-CoA ligase homolog
C:Keywords: carrier protein; cephamycin biosynthesis; phosphopantetheine; phosphoprotein
F:268-758/Domain: acetate-CoA ligase homology <ACLI>
F:786-856/Domain: acyl carrier protein homology <ACPI>
F:1192-1844/Domain: acyl carrier protein homology <ACPI2>
F:1844-1912/Domain: acyl carrier protein homology <ACPI3>
F:2444-2895/Domain: acetate-CoA ligase homology <ACLI3>
F:2912-2980/Domain: acyl carrier protein homology <ACPI3>
F:2920-1896,2444/Binding site: phosphopantetheine (Ser) (covalent) #status predicted

Query Match 4.1%; Score 101.5; DB 1; Length 3649;
Best Local Similarity 20.7%; Pred. No. 93;
Matches 85; Conservative 54; Mismatches 141; Indels 131; Gaps 19;
QY 102 ATLINIKTY--CMGPDHLILEVRDHLRRKQYGGCFILARRMSPALMAG-ASGXGTVDR 158
Db 10H3 AGILNKKSHYVWSLGDYDH-----VQNHRTKRLYLCAOL-----TQALKAGCAADQVT--- 1130

```

```

QY 159 NNGTYLVSTFL-HEGOVSLSLLIHPSEG-----VSALMSARNQGYDRVITGQVNGT 212
Db 1131 -----LHSLQFVWH-----KVLAIGGGNTTIVGTISGHNLPV----- 1166
QY 213 SQVHSECGILTNALNLCQYLDNRDEGFCVPRQHPGACALYHYSKKNVSLYSKQK 272
Db 1167 -GIENSAGLFINT---LPLVHDQAGQNAENAVRIOAANTHNSKS---IVELGSLQS 1220
QY 273 SLFERNVGEVIMEKFNITSKCNLTLSVDLHESGKLQHLQAVLDNRINIQWQYCP 332
Db 1221 GEMKR-----RLFDTLVLE-NYPRILD--DEEELAHQELREK----- 1257
QY 333 LIGSMYSVKEMEYLTRAIDRTGKNTYIVISQHFRRPPIADP--IRRAL----- 383
Db 1258 -----AYDQKQVDPVAVNAREGDELTYLWYNGELFDDDTIDILDVATLTPROVTD 1312
QY 384 -----NWKRAIOHLLNSPDITWITENIR----- 407
Db 1313 IARPVRELDLISPMKARFDSMNETAEEFPADKTLHVFENAEKMPDEIVAVYRENRLT 1372
QY 408 IRMYNDAEERFSDPHGYOYL-----IKQIFQDLSIIDAMDITAY 451
Db 1373 YRELNRANRLAHYRSVVELRPDDVLVALVDKSELMITAIIAKTKGAAV 1423

RESULT 11
T52134
zwille protein [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 20-Oct-2000 #sequence_revision 20-Oct-2000 #text_change 03-Nov-2000
C:Accession: T52134
R:Laux, T.
submitted to the EMBL Data Library, January 1998
A:Reference number: T52134
A:Accession: T52134
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-988 <LAU>
A:Cross-references: EMBL:AJ223508; PIDN:CA11429.1
C:Genetics:
A:Gene: zwille
A:Map position: 5
C:Superfamily: rabbit translation initiation factor eIF-2C

Query Match 4.1%; Score 101; DB 2; Length 988;
Best Local Similarity 20.5%; Pred. No. 16;
Matches 68; Conservative 47; Mismatches 124; Indels 92; Gaps 15;
QY 212 TSGVHSECGILTNALNLCQYLDNRDEGFCVPRQHPGACALYHYSKKNVSLYSKQK 271
Db 391 TQPTRELTFVDEKCTMKSVIETQEMTGFTIOHTLPCLOV---GNQKASYLPMER 446
QY 272 KSLFERNVGEVIMEKFNITSV-----SKCNLTLSVD-----LHESG-KLOH 312
Db 447 CKTVEGQRYRLKLEKQITALLKVTQCORAEQGRNDILRTQHNAAVDQPYAKEGWNISE 506
QY 313 QLVAVDNRINIQWQY-----CYPLISMTYSVKE-----E 345
Db 507 KLASVARTLAPRLKHNHGKQKCLQYQGMKMKKKNKNGMYSMACVNFSSVQF 566
QY 346 YLTRAIDRTGKNTYIVISQHFRRPPIADP--IRRALNVHKAIOHLLNSP----- 397
Db 567 NVARGQNELQDCEV---SGMEFPEVPIRISARPDVYKALKHYHNSMKTKGKE 622
QY 398 -DTMYIKTENIRMYNDAEERFSDPH-GYI-QYLIIDIPQ-----DLVSI----- 441
Db 623 LELLALIPDNNGSLYGDLRKICETELGISQCLTHVYKRSKQYLDVSLKINWKKG 682
QY 442 -----IDA-----WDI-TIAYGTNNVHP 458
Db 683 RNTVIVDAISCRIPLVSDIPTIIFGADVTHP 713

```



```

Db 71 HAFSSDTKEMSPNLT-----SKKKKKKEIRPDIMNSDS- 109
QY 83 IPRPFTVNTTTSATHSATILNPRDTYCGDQLHILLEVDHLGRKKQYCGDFLRAM 142
Db 110 -----SMTSSINKON-----NNOIKSVL-LKENKG-----VKI 136
QY 143 SSPALMAGASQKVTDFNNQTYLVSTFLFWEGVSLILLI-IPSGVSALMSARNGCYR 201
Db 137 TGPG-----NVLSTFLVPIITYIDETKYNMIEKLYEL 169
QY 202 VFTG--QFVNGTSQVHSECGILLTNAELCYLDNRDGEFCYVRQHPHCALTHMYS 259
Db 170 DEFSSTKPKDITTELRTSDDTLMTNPNVGSRODLDKDLVNI-----CA----- 216
QY 260 KKKVSYLSKQKSLPERSNQVEIMEKFTISVSKNTLKSVDLHESGKLQHLAVDLD 319
Db 217 EMTKFKFV-----VYIKNT-----LTLLK-WKYETGVTKNK--VDIR 251
QY 320 RNINIQOKCYPLIGSMYSKMEYLTRAIDRTGKNTVIYISLQHPFPIDVFI 379
Db 252 Q---YKKMLTRPITTIQIHVSSENKD-TLLLR-----SKNYVKTIDIPFCVDMATNCF 303
QY 380 RRAUNVKAIO-HLLRSPDT---MVIKTENIREMYN-----DAERFSDH----- 422
Db 304 SGNINEKCLECTLLVONNDTSECFYVSNQVRENFNOIKAEADENFRNHLDTIN 363
QY 423 -----GYIYLIIKDFODLSVSIIDAMDTITIAVGTNNVPPQHVAGNOINI 469
Db 364 NLKRYIKTNKNGKKELTLEELDFLAKESITDYCKILREIDTNGT-IYNHELGNNDV 422
QY 470 LLYNI 474
Db 423 FNNLI 427

```

RESULT 15

S54069

hypothetical protein YPR045c - yeast (Saccharomyces cerevisiae)

N:Alternate names: hypothetical protein YP9499.03c

C:Species: Saccharomyces cerevisiae

C>Date: 08-Jul-1995 #sequence_revision 19-Oct-1995 #text_change 05-Nov-1999

C:Accession: S54069

R:Badcock, K.: Church, C.M.

submitted to the EMBL Data Library, May 1995

A:Reference number: S54059

A:Accession: S54069

A:Molecule type: DNA

A:Residues: 1-470 <BAD>

A:Cross-references: EMBL:249219; NID:g805025; PTD:g805028; GSPDB:GN00016; MIPS:YPR045c

A:Experimental source: strain AB972

C:Genetics:

A:Gene: MIPS:YPR045c

A:Map position: 16R

Query Match

Best Local Similarity 3.9%; Score 98; DB 2; Length 470;

Matches 80; Conservative 61; Mismatches 129; Indels 138; Gaps 17;

```

QY 86 RPTVNTTTSATSTATIL-----NPRDTYCGDQLHILLEVDHLGRKKQYCG 135
Db 31 RPLDAGSPSLITNNNFIANGQPSLPLPRFLSMNNVNSNPLMVTPLPLGLKKRNNK 90
QY 146 DEL-----RAHMSPALMAGASQKVTDFNNQTYLVSTFLFWEGVSLILLIHPSEGV 191
Db 91 NIKKKLPRVSKKASALSNQVSNV-----MSNSIVGHGAVGSASG 131
QY 192 WSAKNGYD---RVFTGQFVNGTSQVHSECGILLTNAEL----- 229
Db 132 WKVEMGSSDELEKRRKRAERFSGPSAT-----TNSNDNLNEDPANLANISSKSHQYD 184
QY 240 -----CY-----YLDNRDGEFCYVR-----QHPHCALTHMYSKKNKVSYSLSKOEK 272

```

```

Db 185 KLIHVGRQCTLEKSYLRLTSEPNDLRPPNLIQKNCLLMDKXQSKTATYTLDDQPK 244
QY 273 SLFERSNNGVEIMEKFNITSVSKN---TKSVDLHESGKLQHLAVDLDNRNINIQOKY 329
Db 245 SM--RQDLRVQMIENSFTIKYQTHARIALENGDLCFNOCCNRI-MALFENPTIPKSY 301
QY 330 ---CYPLIGSMYSKMEYLTRAIDRTGKNTVIYISLQHPFPIDVFIIRALNV 385
Db 302 SEFCY---SVLYSMLTEDY----- 318
QY 386 HKAIOHLLRSPD--TMVIKTENIREMYN--NDAERFSDHGYIO-YL 428
Db 319 -PSISHLKLKLDGSSSELEDEHVKMIFELSDMKLVGNHYHFPKKNYL 365

```

Search completed: October 12, 2001, 16:04:29
 Job time: 112 sec

Fri Oct 12 16:09:16 2001

us-09-729-454-1.rpt

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

us-09-729-454-1

Run on: October 12, 2001, 16:02:37 ; Search time 34.67 Seconds

(without alignments)
830,585 Million cell updates/sec

Title: US-09-729-454-1

Sequence: 1 MKISMINYKSLALILFLAS.....VHPQHYVGNQINILLNIC 475

Scoring Table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 60623988 residues

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

A_Geneseq-0601.*
1: /SIDSR/gcadata/geneseq/AA1980.DAT.*
2: /SIDSR/gcadata/geneseq/AA1981.DAT.*
3: /SIDSR/gcadata/geneseq/AA1982.DAT.*
4: /SIDSR/gcadata/geneseq/AA1983.DAT.*
5: /SIDSR/gcadata/geneseq/AA1984.DAT.*
6: /SIDSR/gcadata/geneseq/AA1985.DAT.*
7: /SIDSR/gcadata/geneseq/AA1986.DAT.*
8: /SIDSR/gcadata/geneseq/AA1987.DAT.*
9: /SIDSR/gcadata/geneseq/AA1988.DAT.*
10: /SIDSR/gcadata/geneseq/AA1989.DAT.*
11: /SIDSR/gcadata/geneseq/AA1990.DAT.*
12: /SIDSR/gcadata/geneseq/AA1991.DAT.*
13: /SIDSR/gcadata/geneseq/AA1992.DAT.*
14: /SIDSR/gcadata/geneseq/AA1993.DAT.*
15: /SIDSR/gcadata/geneseq/AA1994.DAT.*
16: /SIDSR/gcadata/geneseq/AA1995.DAT.*
17: /SIDSR/gcadata/geneseq/AA1996.DAT.*
18: /SIDSR/gcadata/geneseq/AA1997.DAT.*
19: /SIDSR/gcadata/geneseq/AA1998.DAT.*
20: /SIDSR/gcadata/geneseq/AA1999.DAT.*
21: /SIDSR/gcadata/geneseq/AA2000.DAT.*
22: /SIDSR/gcadata/geneseq/AA2001.DAT.*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	107.5	4.3	4536	20	AAW96826
2	104.5	4.2	3923	20	AAV31237
3	104.5	4.2	4536	19	AAV31262
4	103.5	4.2	954	20	AAV33728
5	101.5	4.1	1560	10	AAW94145
6	100.5	4.0	2721	8	AAW70647
7	100	4.0	935	20	AAV31987
8	100	4.0	1031	19	AAW56102
9	99.5	4.0	705	21	AAH15553
10	99.5	4.0	705	21	AAH15562
11	99	4.0	452	21	AAV70502

12	98.5	4.0	1004	21	AAH18214	Plasmodium falciparum
13	97.5	3.9	623	20	AAV28918	Human regulatory p
14	97.5	3.9	953	20	AAV31141	Human CARD-4L prot
15	97.5	3.9	953	21	AAH15552	Apoptosis related
16	97.5	3.9	953	22	AAH20080	Human CARD-4L (lon
17	97	3.9	906	21	AAH15563	Arabidopsis thaliana
18	97	3.9	983	21	AAH15562	Arabidopsis thaliana
19	97	3.9	988	21	AAH15561	Arabidopsis thaliana
20	97	3.9	1291	19	AAW59912	Plasmodium falciparum
21	96.5	3.9	1802	21	AAH18217	Amino acid sequenc
22	92	3.7	1847	17	AAH92527	Fas antigen #2.. S
23	92	3.7	1847	21	AAH32002	M. jannaschii MJ14
24	92	3.7	1847	21	AAH15631	M. jannaschii MJ14
25	91.5	3.7	1662	13	AAH22561	Mouse mutant c-raf
26	91.5	3.7	1962	12	AAH10563	Mutant protease (K
27	91	3.7	253	21	AAH43066	Human ORF1 ORF2830
28	90.5	3.6	483	11	AAH07445	Secretory signal p
29	90.5	3.6	2325	17	AAW05590	Malze acetyl CoA c
30	90.5	3.6	2325	19	AAW67326	Malze ACCase enzym
31	90.5	3.6	2325	21	AAH14687	Malze acetyl CoA c
32	89.5	3.6	648	13	AAH22560	Mouse mutant c-raf
33	89.5	3.6	648	13	AAH22562	Mouse mutant c-raf
34	89.5	3.6	790	20	AAV37750	Protein involved i
35	89.5	3.6	3418	20	AAV04355	Human BRCA2 (cml2)
36	89	3.6	2383	21	AAH15945	E. coli proliferat
37	88.5	3.6	951	21	AAH31736	Agarase 4-3 protei
38	88.5	3.6	1521	21	AAH39225	Arabidopsis thaliana
39	88.5	3.6	1528	21	AAH39223	Arabidopsis thaliana
40	88.5	3.6	1528	21	AAH39223	Arabidopsis thaliana
41	88.5	3.6	1703	21	AAH36714	Arabidopsis thaliana
42	88.5	3.6	1703	21	AAH36714	Arabidopsis thaliana
43	88.5	3.6	1744	21	AAH36712	Arabidopsis thaliana
44	87.5	3.5	648	13	AAH22563	Mouse mutant c-raf
45	87.5	3.5	648	20	AAH30668	Mutant mouse c-raf

ALIGNMENTS

RESULT 1	AAW96826	standard; protein; 4536 AA.
ID	AAW96826;	
AC	AAW96826;	
XX	22-APR-1999	(first entry)
DE	Amino acid sequence of human apolipoprotein B-100 (apob-100).	
XX	Human apolipoprotein B-100; apob-100; very-low density lipoprotein; VLDL;	
KW	apolipoprotein; binding; in vivo transport; nucleic acid; binding domain;	
KW	nuclear localisation sequence; gene therapy; cancer; cystic fibrosis;	
KW	non-small cell lung carcinoma; diabetes; arteriosclerosis.	
XX	Homo sapiens.	
XX	W09656938-A1.	
XX	17-DEC-1998.	
XX	10-JUN-1998;	98MO-US11927.
XX	14-MAY-1998;	98US-0079030.
PR	13-JUN-1997;	97US-0874807.
XX	(BAYLOR) BAYLOR COLLEGE MEDICINE.	
PA	Guevara JG, Hoogveen RC, Moore JP;	
XX	WPI, 1999-070331/06.	
XX	Composition comprising nucleic acid bound to LDL or VLDL lipoprotein	
PT	- used for delivering nucleic acid to cells for gene therapy and	

PT antisense treatment
 XX
 PS Disclosure; Fig 1A-C; 293pp; English.
 XX
 CC The present sequence represents human apolipoprotein B-100 (apoB-100).
 CC Apob-100 is a major apoprotein component of very-low density
 CC lipoproteins (VLDL), intermediate density lipoprotein (IDL), low
 CC density lipoproteins (LDL) and lipoprotein a. The specification
 CC describes a composition that comprises LDL and apolipoproteins for the
 CC binding and in vivo transport of nucleic acids. Binding domains
 CC (see AAW96878-97) and nuclear localization sequences (see AAW96878-97)
 CC for use in the composition can be derived from the present sequence. The
 CC composition is used to deliver nucleic acids to eukaryotic cells, in
 CC vivo or in vitro, for expressing a therapeutic polypeptide or antisense
 CC molecule (or ribozyme). Specifically they are used for gene therapy of
 CC cancers (particularly non-small cell lung carcinoma), diabetes, cystic
 CC fibrosis and arteriosclerosis.
 XX
 SQ Sequence 4536 AA;

Query Match 4.3%; Score 107.5; DB 20; Length 4536;
 Best Local Similarity 20.7%; Pred. No. 5.1;
 Matches 90; Conservative 71; Mismatches 169; Indels 105; Gaps 22;

OY 56 KTPILSLKPLTETE-LRIKELIEKLDQOIPPPPTHV---NTTTSATHS-TATILNPRD 109
 DB 1998 kypilseplnildalemrdaevk-----pqetlvaatykydkngdvhslnlpfetlq 2051
 OY 110 TYCRGDQHLHLEVRDHLGRRKOYGD-FLRARRSSPALMAGASGVKVDNFNGTYLVST 168
 DB 2052 eyfernrgtllvvenvqgnlkhnlndqfvk-----yralqklpqqan-dynlsln 2103
 OY 169 LFMGGVSLSLLIHPSEGSVALMS-----ARNGCYDRVTF 204
 DB 2104 ---wergvs-----hakekitalctkyrttendqlaldaklnfneklslqlytq--mldq 2153
 OY 205 TGGFVNGTSQVHSECGILNTNAELCOYLNRDQEGFVCVRPQHPMCAALTHMYSK---- 260
 DB 2154 fdqyldsydlhdhklalaneldeleklsidsh--yhlrvnlvktlhdhlnfendf 2211
 OY 261 NKKVSYLSKOEKSLFERSNVGVEIMEKNITSVSKNTLKSVDL-HSGKL-OHOLAVDL 318
 DB 2212 nksqstaswvgnvdkyqirtdqeklqk-----rhqndldqhlqagklkqhldeadv 2267
 OY 319 DRNINIQOKTCYPLIGSMTYSVKEMEYLTRAIDRTGGEKNTVIVISLGOHFRPPPT--- 375
 DB 2268 -----tvllldlqgt-----elstferl----ndvl-----ehvkhftvlnl 2298
 OY 376 -DVFLRALNVHKAIOHLILRSPT-----MVIKTENIREMYNDARSPDRHGTYOVL 429
 DB 2299 gdfveaeklnafakvhehleryevdqqlvmdkvlvelthqylkkelqklslsnvlgvsk 2358
 OY 430 IKDIFODLSVSIIDA 444
 DB 2359 lkydfeklvqfidda 2373

RESULT 2
 AAY31237
 ID AAY31237 standard; Protein; 3923 AA.
 XX
 AC AAY31237;
 XX
 DT 08-NOV-1999 (first entry)
 XX
 DE Human Apo B protein fragment.
 XX
 KM Apo B; hyperlipidemia; human; treatment; hepatocyte; apoprotein;
 KM Apo A1; low density lipoprotein; LDL; blood; therapy; arteriosclerosis;
 KM high density lipoprotein; HDL; cholesterol; coronary heart disease;
 KM Alzheimer's disease; hypobetalipoproteinemia; dysbetalipoproteinemia.
 XX

OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT 1..3923 Protein
 FT /label= Apo B
 FT /note="Partial sequence, no start codon given"
 XX
 PN M09940789-A1.
 XX
 PD 19-AUG-1999.
 XX
 PF 28-AUG-1998; 98MD-US17908.
 XX
 FR 30-JUN-1998; 98US-0108006.
 PR 12-FEB-1998; 98US-0074497.
 XX
 PA (MING) UNIV MINNESOTA.
 PA (YESH) UNIV YESHIVA EINSTEIN COLLEGE.
 XX
 PI Bandypadhyay PT, Kren BT, Roy-Chowdhury J, Steer CJ.
 XX
 DR WPI: 1999-527333/44.
 DR N-PSDB: AA209525.
 XX
 PT Mutating apolipoprotein genes in hepatocytes to control cholesterol
 PT levels, e.g. for treating or preventing hyperlipidemia, particularly
 PT arteriosclerosis
 XX
 PS Claim 4: Page 75-83; 106pp; English.
 XX
 CC This invention describes a novel method for the genetic treatment of
 CC hyperlipidemia by altering genes, in hepatocytes, for apoprotein (apo) B,
 CC E or A1. Low density lipoprotein (LDL) levels in the blood are reduced by
 CC altering an apo B gene (1) in a hepatocyte. The invention describes a
 CC method for the therapeutic and/or prophylactic method involving altering
 CC an apo E gene in hepatocytes by introducing the mutations Arg12Cys,
 CC Arg158Gys or Cys158Arg and a method for ameliorating arteriosclerosis by
 CC lowering the apo A1 gene in a hepatocyte so that the altered protein can
 CC dimertize. Altering expression of apo genes regulates levels of high and
 CC low density lipoprotein cholesterol. Altering expression of apo B, E and
 CC A1 genes is used to treat or prevent arteriosclerosis, coronary heart
 CC disease, Alzheimer's disease, hypobetalipoproteinemia, and
 CC dysbetalipoproteinemia. This sequence represents a treatment of the human
 CC Apo B protein described in the method of the invention.
 XX
 SQ Sequence 3923 AA;

Query Match 4.2%; Score 104.5; DB 20; Length 3923;
 Best Local Similarity 20.5%; Pred. No. 7.8;
 Matches 89; Conservative 72; Mismatches 169; Indels 105; Gaps 22;

OY 56 KTPILSLKPLTETE-LRIKELIEKLDQOIPPPPTHV---NTTTSATHS-TATILNPRD 109
 DB 1385 kypilseplnildalemrdaevk-----pqetlvaatykydkngdvhslnlpfetlq 1438
 OY 110 TYCRGDQHLHLEVRDHLGRRKOYGD-FLRARRSSPALMAGASGVKVDNFNGTYLVST 168
 DB 1439 eyfernrgtllvvenvqgnlkhnlndqfvk-----yralqklpqqan-dynlsln 1490
 OY 169 LFMGGVSLSLLIHPSEGSVALMS-----ARNGCYDRVTF 204
 DB 1491 ---wergvs-----hakekitalctkyrttendqlaldaklnfneklslqlytq--mldq 1540
 OY 205 TGGFVNGTSQVHSECGILNTNAELCOYLNRDQEGFVCVRPQHPMCAALTHMYSK---- 260
 DB 1541 fdqyldsydlhdhklalaneldeleklsidsh--yhlrvnlvktlhdhlnfendf 1598
 OY 261 NKKVSYLSKOEKSLFERSNVGVEIMEKNITSVSKNTLKSVDL-HSGKL-OHOLAVDL 318
 DB 1599 nksqstaswvgnvdkyqirtdqeklqk-----rhqndldqhlqagklkqhldeadv 1654
 OY 319 DRNINIQOKTCYPLIGSMTYSVKEMEYLTRAIDRTGGEKNTVIVISLGOHFRPPPT--- 375

Db 1655 -----rvllldqgl-----tsterl-----ndvl-----ehvkhvlnll 1685
 Qy 176 -DVFIRALNVHKAIOHLLRSPDT-----WVITKENIRBMNDARESDHGVIQYLI 429
 Db 1686 gdfveaekinafrakvheilerjevqqlqylmcklvelahqyklketlqlsnvlqgvk 1745
 Qy 440 IKDIFODLSVSIIDA 444
 Db 1746 ikdyfeklvglidda 1760

RESULT 3

AAW41262
 16 AAW41262 standard; peptide; 4536 AA.
 AAW41262;
 D0 19-MAY-1998 (first entry)
 XX
 De Apolipoprotein B-100.
 XX
 KW Anti-coagulant; apolipoprotein B-100; apob-100; metastatic spread;
 KW thromboplasmin-mediated process; cancer; inhibitor; blood coagulation;
 KW angiogenesis; cellular differentiation; apoptosis; KRAD-14;
 KW prothrombinase complex.
 OS Homo sapiens.
 XX
 PN W09743311-A1.
 XX
 P0 20-NOV-1997.
 XX
 PF 09-MAY-1997; 97W0-GH01255.
 XX
 PR 09-MAY-1996; 96GB-0009702.
 XX
 PA (UNLO) ROYAL FREE HOSPITAL SCHOOL MED.
 XX
 PI Bruckdorfer KR, Ettelate C;
 XX
 DR MPI: 1998-008798/01.
 XX
 PT Peptide fragments of apolipoprotein B-100 with anticoagulant
 PT activity - used for treating or preventing coagulation, inhibiting
 PI angiogenesis, cell differentiation and apoptosis
 XX
 PS Disclosure: Page 42-47; 60pp; English.

XX This sequence represents the human apolipoprotein B-100 (apob-100).
 CC Fragments of this sequence can be used in the peptide of the invention,
 CC which has the formula (I), or their variants with one or more internal
 CC deletions, insertions or substitutions, while retaining anti-coagulant
 CC properties of apolipoprotein B-100 (apob-100).
 CC Z1 KACVXIKKKKRRHS-A2-T-22 (1): X1 - S or Y; X2 - T or I;
 CC Z2 - the N terminus of the peptide, or 1-47 amino acids (aa);
 CC Z3 - the C terminus of the peptide, a terminal amide group or 1-77 aa.
 CC Compositions containing the peptide, are used for simultaneous, separate
 CC or sequential treatment of cancer, particularly to prevent metastatic
 CC spread. They are also used to inhibit thromboplasmin-mediated processes,
 CC specifically to prevent or reduce blood coagulation (e.g. during or after
 CC surgery or in cases of heart attack, stroke etc.) and to inhibit
 CC angiogenesis, cellular differentiation or apoptosis. KRAD-14, which is
 CC active as such or as part of a 98-aa peptide, inhibits activation of the
 CC prothrombinase complex; and prevents activation of factor VII on the
 CC surface of thromboplasmin and of platelets by thrombin. It binds to the
 CC residues 58-66 of thromboplasmin. Since (I) are much smaller than
 CC apob-100, they act more quickly.

XX Sequence: 4536 AA;
 SU

Query Match

4.24; Score 104.5; DB 19; Length 4536;

Best Local Similarity 20.54; Pred. NO. 10;
 Matches 89; Conservative 72; Mismatches 169; Indels 105; Gaps 22;
 Qy 56 KTLPLSLKPLTELE-LRIKEILEKLDQIIPPPFTVH-----NTTSATHS-TAILNPRO 109
 Db 1998 kvplllseplnldldemravck-----pqetfvafrkkykngdvhslnlpfctq 2051
 Qy 110 TYCKGQDLHTLEVDHIGRRKQYGD-FLARMSPPALMAGSGKVTDFNNGTYLVSTP 168
 Db 2052 eyfermqclivlenwqnlklnldglvrk-----yralgklpqgan-dylnsfn 2103
 Qy 169 LFMFGVSLSLLLHPSECVSALMS-----ANNGCYDHYR 204
 Db 2104 --wergvs-----hakekitaltkyritendqlaldaklnfeklsqgly--mlq 2153
 Qy 205 TGGFVNGTSQVHSECGILMTNMLELCQYIDRDEGTCVPOHMPCALTHMYSK----- 260
 Db 2154 fdqylkdsydlndlklaiaaidelieklsidh--yhlrvnlvklndhlhlfenldf 2211
 Qy 261 NKXVSYLSKQKSLFERSNVGEIMEKENTISVSKNTLKSVDL-HESGKL-OHOLAVDL 318
 Db 2212 nksqsslaswlnqndtkyqlrlqqlqkqlk-----rhqnlldqlhlaqklkqhlleaiv 2267
 Qy 319 DRNINIQWQKCYPLIGSMYSKEMEYLTADRTGCKNTYIVISIGHFPPPI--- 375
 Db 2268 -----rvllldqgl-----tsterl-----ndvl-----ehvkhvlnll 2298
 Qy 376 -DVFIRALNVHKAIOHLLRSPDT-----WVITKENIRBMNDARESDHGVIQYLI 429
 Db 2299 gdfveaekinafrakvheilerjevqqlqylmcklvelahqyklketlqlsnvlqgvk 2358
 Qy 430 IKDIFODLSVSIIDA 444
 Db 2359 ikdyfeklvglidda 2373

RESULT 4

AAV33728
 ID AAV33728 standard; Protein; 954 AA.
 XX
 AC AAV33728;
 XX
 D0 09-NOV-1999 (first entry)
 XX
 DE Photornhabus luminescens 954 amino acid insecticidal toxin.
 XX
 KW Symbiotic bacterium; nematode; insect; larva; toxin; insecticide.

XX Photornhabus luminescens.
 OS
 PN W09942589-A2.
 XX
 PD 26-AUG-1999.
 XX
 PF 18-FEB-1999; 99W0-EP01015.
 XX
 PR 20-JAN-1999; 990S-0116439.
 PR 20-FEB-1998; 980S-0027080.
 XX
 PA (NOVS) NOVARTIS AG.
 PA (NOVS) NOVARTIS-ERFINDUNGEN VERM GES MBH.
 PI Anderson AR, Chen JS, Dunn MM, Hart HP, Kramer VC;
 PI Morgan MK, Warren GW;
 XX
 DR MPI: 1999-527479/44.
 DR N-PSDB: AAZ06831.
 XX
 PT New nucleic acid from Photornhabus luminescens encoding insecticidal
 PT toxins, used for making resistant transgenic plants
 XX
 PS Claim 26; Page 130-133; 148pp; English.

CC This sequence represents a 954 amino acid insecticidal
 CC toxin from *Photobacterium luminescens*. It is one of three insecticidal
 CC toxins (AAV3728-Y3730) encoded by open reading frames (ORFs)
 CC in a 38kb fragment of *P. luminescens* DNA (AAZ06831). This sequence
 CC is encoded by ORF5. *P. luminescens* is a member of the
 CC Enterobacteriaceae family and is a symbiotic bacterium of
 CC nematodes of the genus *Heterorhabdus*. The nematodes colonize insect
 CC larvae, kill them, and their offspring feed on the dead larvae. However,
 CC the insecticidal agents are produced by *P. luminescens* rather than the
 CC nematodes. The toxins have activity against lepidopteran insects such as
 CC Cabbage Looper (*Trichoplusia ni*), European Corn Borer (*Ostrinia
 CC nubilalis*) and Fall Armyworm (*Spodoptera frugiperda*), and also against
 CC Coleopteran insects (e.g., Colorado Potato Beetle, *Leptinotarsa
 CC decemlineata*). In addition the toxins are active against strains
 CC resistant to known insecticides. The DNA sequence can be used to generate
 CC transgenic plants of various species that are resistant to economically
 CC important insect pests and also for recombinant production of the toxins
 CC for use as insecticides.

CC Sequence 954 AA:

Query Match 4.28; Score 103.5; DB 20; Length 954;
 Best Local Similarity 20.18; Pred. No. 0.93;
 Matches 96; Conservative 80; Mismatches 158; Indels 143; Gaps 24;

QY 33 VMSALNLSISLHVMNNSKSLFPTPLI--SLKPLTERLRIKEIEIKDDQIPRPTH 90
 DB 325 VLSIRNDAESTRFHNN--qkwapenlytdslyqjlastgemanlyggsnqdp----- 376
 QY 91 VNTTSTATSTATILNPRDTCRGDDHLHLLEVRDHLGRKKQYGGFPLARMSSPALMAG 150
 DB 377 -SLTPEDMNYCYNYTCTYDTR-----ggnlLkLqhsstacqnn 415
 QY 151 ASGKVTDFNNGTYLVSTFLFMEGVSLSLILHPSEGVSLMSARNOGDRVFTGQFVN 210
 DB 416 YLTLITVMSNRNAV-----Lstltedpaq-vdalfda--ggnqntllsgnln 461
 QY 211 GTSQVHSEGGILLNMAELQYLD--NRQGEF-----YCVAPQHMPCAA 253
 DB 462 -----wntlrgel-qhwclvkrckgandrewygyssdgrllklineq----- 503
 QY 254 LTHWYSKNKKVSYLSKQEKSLFERSNVGVEIMEKENTISVSKNCTLSKSLFERS- 311
 DB 504 -tsnasqtrlllylpolelrltqnsfitedlg--Yltvgeagraqrvylhwdsqged 559
 QY 312 ---HOL-----AVDLDRNINIQOKCYCPPLIGSMRTYSV-KEMEYLTRAIDRTGG 356
 DB 560 IdmnglryeydnlIgsqglldskgeIseeeypyysgtalwatrlkrtasaykllrys 619
 QY 357 EKNTVIVISLG-QHFRFP-----PIDVFRRLNVHKAIOH--LLRSPDTMVI 402
 DB 620 erdatgllyyyryyqpwwgfwlsadpaqtdv-----glnlyymvrnmpvllldpdl 672
 QY 403 IKTENIREMYNDAREHFSDFHGYOYLIIKDFOLSLSIIDADITIAVGNNHNP 459
 DB 673 -----mptlaetla-----alqknkvadapsrlnatnval-----nlrpp 708

RESULT 5

AAAP94145

ID AAP94145 standard; protein: 1560 AA.

XX AAP94145;

XX 07-JUN-1990 (first entry)

XX S. cremoris protease.

XX Lactic acid bacteria; cheese; Streptococcus cremoris SK112;

XX protease; psk112; chymosin; prochymosin.

XX NL8701378-A.

XX PD 02-JAN-1989.
 XX PF 12-JUN-1987; 87NL-0001378.
 XX PR 12-JUN-1987; 87NL-0001378.
 XX PA (NEZU-) NEDERL INS ZUIVELON.
 XX PT Simons AFM, De Vos WM;
 XX DR WPI: 1989-030097/04.
 XX NR P-PSDB; AAN91159.
 XX PT DNA fragment having region specific for lactic acid bacteria -
 XX PT is contained in plasmid in microorganism used in prodn. of
 XX PT protein and food prodn. eg cheese.
 XX PS Claim 6: fig 7a; 43pp; Dutch.

XX The DNA encoding prochymosin can be cloned into a plasmid (esp from
 CC S. cremoris SK112) and used to produce large amounts of the protein by
 CC recombinant DNA techniques. The protein synthesis is driven by the
 CC regulatory region of this protease. This could overcome the shortage
 CC of prochymosin due to a shortage of calf stomachs and increasing cheese
 CC prodn. Prochymosin is also used in prodn. of yoghurt, butter and
 CC buttermilk.
 CC See also AAP94144-P94146.

XX Sequence 1560 AA:

Query Match 4.18; Score 101.5; LB 10; Length 1560;
 Best Local Similarity 21.18; Pred. No. 3.3; 177; Indels 131; Gaps 21
 Matches 96; Conservative 52; Mismatches 177; Indels 131; Gaps 21

QY 58 PLISLKLPTETELRIK-----EIEKIDQDIPRPFTHVNTTSTATILNPK 106
 DB 714 pavellkdfstidkfkilfnttletlyqnds-----ntdnavytsatdpn-- 761
 QY 109 DTVCRGDDQHILLEVRDHLGRKKQYGGFPLARMSSPALMASGVYDFPNNGYLVST 168
 DB 762 -----sgvlydkkldgaalaaqsn--ltvpqkdaalelt 794
 QY 169 LFWEGVSLSLILHPSEGVSLMSARNOGDRVFTGQFVNCTSVY-HSECG----- 420
 DB 795 L-----slpkslnlpymgffgqwn-----dgklydslogllyspaqndutp 837
 QY 221 LIINTNAELCOYLON--ROQEGPYCVRQHMPCALTHMYSKKNKVSLSKQEKSLFERS 276
 DB 838 llnkntg-tlyygmwtadadngqvddq-----aaisdksua; yndisnky; llnns 871
 QY 279 NQGVETME-----KENTISVSKNCTLSKSLVSLHESKRLQHOLAV-----DIDPFINI 927
 DB 892 nqvvdlldggnkvctllssstnrkkyynaahsqgylynapawdqtydqg dnyktaall 951
 QY 328 KYCYPILGSMYTSVSKEMEYLTRAIDRTGGEKNTVIVISLGQHFRFPEDVFI-----RR 961
 DB 952 -----gslytyrlysgyp-----eggdktqfydv-----pklksktrp- 1004
 QY 382 ALNVHKAIOHLLRSPDTWYIKTENIR-ENYNDAREHFSHFHGYOYLIIKDFOLSL 437
 DB 990 vldvpfkldskaprvhvalsaakntgktyylaekadfgslaktkvtoivetvl 1049
 QY 438 SVSIIDAMDITIAVGNNHNP-----QHVVGQGIN 468
 DB 1050 datfcdag--ltadgytkletplseagqalngdn 1083

RESULT 6

AAAP70647

ID AAP70647 standard; protein: 2721 AA.

```

AC  AAP70647:
XX
XX  17-APR-1991 (first entry)
XX
XX  Sequence of N-terminal apolipoprotein B (apob).
XX
XX  Lipid-binding peptide; apolipoprotein purification.
XX
XX  Homo sapiens.
XX
XX  Key Location/Qualifiers
XX  Feature 87..167
XX  Protein /label=Signal 168..8247
XX
XX  W08702062-A.
XX
XX  09-APR-1987.
XX
XX  02-OCT-1986; 86MO-US02075.
XX
XX  04-DEC-1985; 85US-0804692.
XX  04-OCT-1985; 85US-0784418.
XX  27-FEB-1986; 86US-0834300.
XX
XX  (BIOT-) BIOTECHN RES PARTN.
XX
XX  Procter AA, Vigne JL, Mallory JB, Talmadge KD.
XX
XX  WP1: 1987-108703/15.
XX  N-PSDB: AAN70996.
XX
XX  Purified lipid-binding peptide prodn. - by expression of
XX  recombinant system and adding lipid to obcd. peptide mixc. to
XX  form lipo-peptide complex
XX
XX  Claim 50; pp73-85; 123pp; English.
XX
XX  Apob is a lipid-binding peptide adapted to bind to phospholipids at
XX  one or more amphipathic alpha-helical peptide regions. Apob gene is
XX  pref. expressed in Chinese hamster ovary (CHO) cells and the gene is
XX  introduced in an expression vector having a regulatable promoter
XX  derived from the human metallothionein II gene.
XX
XX  Sequence: 2721 AA:

Query Match 4.0%; Score 100.5; DB 8; Length 2721;
Best Local Similarity 20.5%; Pred. No. 10;
Matches 89; Conservative 71; Mismatches 170; Indels 105; Gaps 22;

```

```

DB  2295 -----TVLLDQIGT-----tisferl----ndvl-----ehvkhfvlnpy 2325
QY  376 -DVFIRALNVHKAIOHLLRSPT-----MWIKTENIRRMVDARFSDPHGYIOLLI 429
DB  2326 wdfwaeakinafrakvhehlerjevqdqhlqvlnmklvelnqykhketqklsnliqyvk 2385
QY  430 IKDIFODLSVSIIDA 444
DB  2386 lkdyfeklvqfidda 2400

RESULT 7
AAV31987
ID  AAV31987 standard; Protein; 935 AA.
XX
XX  AAV31987;
XX
XX  21-DEC-1999 (first entry)
XX
XX  Alpha-ketoglutarate dehydrogenase E1 subunit.
XX
XX  Glutamic acid; alpha-ketoglutarate dehydrogenase; AKGDH; sucAB;
XX  sucA.
XX
XX  Enterobacter agglomerans.
XX
XX  EP952221-AA.
XX
XX  27-OCT-1999.
XX
XX  17-MAR-1999; 99EP-0105508.
XX
XX  18-MAR-1998; 98JP-0069068.
XX  19-OCT-1998; 98JP-0297129.
XX
XX  (AJIN ) AJINOMOTO KK.
XX
XX  Izut H, Ono E, Matsui K, Moriya M, Ito H, Hara Y;
XX
XX  WP1: 1999-582803/50.
XX  N-PSDB: AA220083.
XX
XX  Novel bacteria used for production of L-glutamic acid -
XX
XX  Example 3; Page 19-22; 32pp; English.
XX
XX  The present sequence represents the Enterobacter agglomerans
XX  AJ13355 (FERM P-16644) Alpha-ketoglutarate dehydrogenase (AKGDH)
XX  E1 subunit. The sequence was deduced from the sucA gene
XX  identified in a cloned fragment (see AA220083) of AJ13355 DNA.
XX  This cloned fragment also included the sub open reading frame
XX  encoding the AKGDH E2 subunit (see AAV31988). The invention provides
XX  Enterobacter and Serratia microorganisms that have an improved
XX  ability to produce L-glutamic acid as a result of an increase in
XX  activity of an enzyme catalysing a reaction for L-glutamic acid
XX  biosynthesis and/or a decrease or deficiency in an activity of an
XX  enzyme that catalyses a reaction branching away from a pathway for
XX  L-glutamic acid biosynthesis, such as AKGDH. An E. agglomerans
XX  strain deficient in AKGDH was produced using the sucAB gene by
XX  homologous recombination.
XX
XX  Sequence 935 AA:

Query Match 4.0%; Score 100; DB 20; Length 935;
Best Local Similarity 20.3%; Pred. No. 2;
Matches 86; Conservative 42; Mismatches 136; Indels 160; Gaps 23;

```


Db 186 kvkkyvelnhkhllekintrelsmvyeisakhyffdhemi-----yvlwkl 440
 QY 431 KUIFODLSSTI 442
 Db 441 twlftedlwsll 452

RESULT 9
 AAB15559
 ID AAB15559 standard; Protein: 705 AA.

XX AAB15559;
 DT 28-FEB-2001 (first entry)
 DE Apoptosis related protein encoded by gene 2 clone HDBPM68.

KW Cytostatic; antiinflammatory; immunosuppressive; antisclerotic; cardiant;
 KW virucidal; anti-AIDS; vasotrophic; anti-ischaemic; antiparkinsonian;
 KW anti-Alzheimer; gene therapy; human; apoptosis; fusion protein; cancer;
 KW colon; breast; prostate; melanoma; lymphoma; inflammation; herpes;
 KW autoimmune disorder; multiple sclerosis; viral infection.

XX Homo sapiens.
 XX M0200056752-A2.

XX 28-SEP-2000.

XX 15-MAR-2000; 2000MO-US06642.

XX 24-MAR-1999; 99US-0126018.

XX 17-JUN-1999; 99US-0139638.

XX 18-AUG-1999; 99US-0149449.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Ruben SM, NI J, Young PA;

XX WP1: 2000-587660/55.

XX N-PSDR: AAB155798.

XX Nucleic acids encoding human apoptosis associated protein, useful for

XX the prevention, treatment and diagnosis of e.g. Alzheimer's and

XX Parkinson's disease, inflammation and ischaemic injury -

XX Claim 11: Page 265-267; 273pp; English.

XX The invention relates to the isolation of genes encoding a human

XX apoptosis-related proteins. The nucleotide sequences AAB155790-A95798

XX encode the human apoptosis-related proteins AAB15551-B15559. The genes

XX can be used to generate fusion proteins by linking to the gene for the

XX human immunoglobulin G Fc (196 Fc) portion (AAB155799) for increasing the

XX stability of the fusion protein as compared to the human protein only.

XX The gene and encoded protein may be used in the prevention, treatment

XX and diagnosis of diseases associated with inappropriate apoptosis

XX associated protein expression, e.g. cancer (e.g. colon, breast and

XX prostate cancer, melanoma and lymphomas), inflammation, autoimmune

XX disorders (e.g. multiple sclerosis) and viral infections (e.g. herpes)).

XX Sequence 705 AA:

Query Match 4.0%; Score 99.5; DB 21; Length 705;

Best Local Similarity 20.5%; Pred. No. 1.4;

Matches 96; Conservative 58; Mismatches 154; Indels 161; Gaps 23;

QY 18 IASMTITVFQNSTKXWASALNLSLHYMNNSKSLFPTPTLISKPLTELELRIFEITE 77
 Db 144 IICWILIFCTG-----H-----RADEGSPQIPACTMCLIDVFI--VTE 182
 QY 78 KLDUUIPPROPTHVNTTTSATHTATILNPRDTYG-----RGDQLHILLEVRHLG-- 128

Db 183 vhlrmgpsalvgtrtrs---pyeclhagritdlsclsgvqhgmekslfvtltgeevxas 238
 QY 129 -----RRKQYGGDFLRKRRSSPALMAAGSGKVTDFNNKGYLVSTLTW---EQVSLSL 181
 Db 239 glqetdmqyg--flra---lpeigpqdgqxyefhllqafatfflvdrrvgtel 293
 QY 182 IHPSEGVSAIWSARNQY-----DRVIFPGGVNGTSQVHSE 218
 Db 294 rffgwmppagaatcscypflpfcqlsgsparedlfxkndhfgtfnlfi----- 344
 QY 219 GGLLTNNMELCOYLDNRDQGFYCPVAPQHMCAAL-----THMYSKNKVSYLSKQ 270
 Db 345 cglstskakqkllfhl-----vpaalrrkklkalwalfasll--gylksl 387
 QY 271 EKSLFERSNVGVEIMEKF-----NTISVSKCNTLK----- 300
 Db 388 prvgvesfn-gqamptflmrlcrltsgsqvqglaaqgicanyklklycnacacasa 446
 QY 301 -SYDLHESGKLQHOLAVDLRN-----INIQMKYCYPLIGSMYVSKENEYLTTRADRTG 355
 Db 447 lsfvlhfrp---rlaldlnmndygvrelqpcfsrlvrlsvnqf-----tdg 495
 QY 356 GER-----NTVIVISLQGFHFRFPFDV---FIRRALNWKAIQHL 394
 Db 496 gkvklseeltkyklvtyglgnqf--tdvgaryvklldcekgjthlkl 543

RESULT 10
 AAB15562
 ID AAB15562 standard; peptide; 705 AA.

XX AAB15562;
 DT 28-FEB-2001 (first entry)

XX Fragment of apoptosis related protein encoded by gene 2 clone HDBPM68.

XX Cytostatic; antiinflammatory; immunosuppressive; antisclerotic; cardiant;

XX virucidal; anti-AIDS; vasotrophic; anti-ischaemic; antiparkinsonian;

XX anti-Alzheimer; gene therapy; human; apoptosis; fusion protein; cancer;

XX colon; breast; prostate; melanoma; lymphoma; inflammation; herpes;

XX autoimmune disorder; multiple sclerosis; viral infection.

XX Homo sapiens.

XX M0200056752-A2.

XX 28-SEP-2000.

XX 15-MAR-2000; 2000MO-US06642.

XX 24-MAR-1999; 99US-0126018.

XX 17-JUN-1999; 99US-0139638.

XX 18-AUG-1999; 99US-0149449.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Ruben SM, NI J, Young PA;

XX WP1: 2000-587660/55.

XX Nucleic acids encoding human apoptosis associated protein, useful for

XX the prevention, treatment and diagnosis of e.g. Alzheimer's and

XX Parkinson's disease, inflammation and ischaemic injury -

XX Disclosure: Page 20-21; 273pp; English.

XX The invention relates to the isolation of genes encoding a human

XX apoptosis-related proteins. The nucleotide sequences AAB155790-A95798

XX encode the human apoptosis-related proteins AAB15551-B15559. This

XX sequence represents a fragment of the protein encoded by the gene given

XX in the descriptor line. The genes can be used to generate fusion

PA (HOFF/) HOFFMAN S.
PA (GARD/) GARDNER M.
PA (GARD/) GARDNER M.
PA (VENT/) VENTER J C.
PA
PI Hillman S., Garucci D., Gardner M., Venter JC;
XX
XX WPI: 2000-365347/31.
XX

PT Proteins encoded by chromosome 2 of the human malarial parasite,
PT plasmodium falciparum, useful as antimalarial vaccines and in the
PT diagnosis of P.falciparum infection .
XX
XX
XX

PS Disclosure: Page 165-168: 577pp: English.

XX The present invention describes proteins and their fragments (I) encoded
CC by chromosome 2 of the human malarial parasite, Plasmodium falciparum.
CC Also described are: (1) nucleotide sequences (II) encoding (I); and (2)
CC vaccines against P. falciparum infection comprising (I) or (II).
CC (I) and (II) are useful for the development of vaccines against
CC P. falciparum infection. (I) and polyclonal antisera or a monoclonal
CC antibody raised to immunogens comprising the sequences of (I), are
CC useful in the detection of infection with P. falciparum. Furthermore,
CC (I) especially when they are rifins or secreted or membrane proteins)
CC can aid the identification of drugs to treat or prevent P. falciparum
CC infection, or they can be used to identify drug resistance in
CC P. falciparum. Sequencing of the Plasmodium chromosome 2 and the
CC subsequent identification of the proteins encoded by it will help to expand
CC our understanding of parasite biology, a process hampered by the
CC complexity of the parasite lifecycle, and provide new targets for
CC vaccine and drug development. Parasite resistance to drugs and mosquito
CC resistance to insecticides have led to a resurgence of malaria in many
CC parts of the world, and there is a pressing need for vaccines and new
CC drugs. AAA70078 to AAA70267 and AAB16144 to AAB16352 represent nucleotide
CC and protein sequences given in the present invention, but which are not
CC specifically mentioned within the specification.
XX
XX Sequence 1004 AA:

Query Match 4.0%; Score 98.5; DB 21; Length 1004;
Best Local Similarity 17.5%; Pred No. 3.1;
Matches 85; Conservative 88; Mismatches 151; Indels 161; Gaps 22;

UY 24 IFTYVNSTKWSALNLSISLHYNNSTSLPPTPLSLKPLTELEIKETIKELDQ 82
TH 1 hasissdlkewspnlt-----skkkkkkketlpkdlmsns- 109
UY 83 IPRRFTHVNTTSATHTATILNPRDYCRGDOLHILEVRDHLGRKQYGDPLRAB 142
DB 110 -----ntsslnkgn-----nqdksvl-ikeny-----vkl 136
UY 143 SSIPLAAGASCKVTFPNNGTVLVSFTLFMEGQVSLILLI HPSEGVSAIWSARNQYDR 201
DB 117 tgcpc-----nymlsfllyphylydvekylnlekyel 169
UY 202 VFTG--QFVNGTSQVHSEGLINTNAELCOYLDNRQECFCVPRQMPAALTHYS 259
DB 170 delstskikfkktteltsdltlmnnhnygsrtdlkdjyni-----ca----- 216
UY 260 KKKVYLSKQKSLFERSNNGVEIEMKFNNTSVSKCTLKASVDLHESKLOHOLAVLD 319
DB 217 enktktiv-----vylkdn-----llk-kyvvetgvtlnk--vdllr 251
UY 320 RNINIQMKVYPLIGSMYTSVKEMEYLRADRTGKGNVYVLSIGQHPFPIDVFI 379
DB 252 q---ymkeltlptltqlghsvsensk-d-llle-----sknyvlkldlpctcdmatncfl 303
UY 380 RRAINVHKAID-HLLRSPT--WVITTEIREMKN-----DAEFSDFH----- 422
DB 364 sqnlueklectllvgnndtseectlysnvdeventnqkkaeeddnfrnyhltdln 363
UY 423 -----GYOYLILIKDIFODLSVSLTDAMDITAYGNTNVHPPQHVGNOINI 469

DB 364 nllkriyknkgkkelitlleednfkseltdyckllreldngt-lynhelgnnydv 422
UY 470 LNNYI 474
DB 423 fnnll 427

RESULT 13
AAV28918
ID AAV28918 standard; Protein: 623 AA.
XX
AC AAV28918;
XX

DT 21-SEP-1999 (first entry)
XX
DE Human regulatory protein HRP-4.
XX

KM Regulatory protein: HRP; human; cell proliferation; immune response;
KM cancer; adenocarcinoma; leukemia; lymphoma; melanoma; myeloma; sarcoma;
KM tetracarcinoma; Addison's disease; adult respiratory distress syndrome;
KM allergy; anemia; asthma; atherosclerosis; bronchitis; cholecystitis;
KM Crohn's disease; ulcerative colitis; atopic dermatitis; dermatomyositis;
KM diabetes mellitus; emphysema; atrophic gastritis; glomerulonephritis;
KM gout; Grave's disease; hyperlipidemia; irritable bowel syndrome; AIDS;
KM lupus erythematosus; multiple sclerosis; myasthenia gravis; Intarction;
KM osteoarthritis; osteoporosis; pancreatitis; polymyositis; hemodialysis;
KM arthritis; scleroderma; Sjogren's syndrome; and autoimmune thyroiditis;
KM infection; trauma.

OS Homo sapiens.
XX
XX WO933870-A2.
PN 08-JUL-1999.
XX
XX 22-DEC-1998; 98MO-US27471.
PF
XX
XX 31-DEC-1997; 97US-0001403.
PR

(INCY-) INCYTE PHARM INC.
PI Au-Young J, Bandman O, Corley NC, Guegler KJ, Hillman JL;
PI Lal P, Shah P, Tang YT, Yoe H;
XX
XX WPI: 1999-430229/36.
DR
XX N-PSDB: AAX09288.

New human regulatory proteins, useful for diagnosing, preventing and
treating disorders associated with expression of regulatory proteins
Claim 1; Page 71-72; 89pp: English.

The invention provides novel human regulatory proteins (AAV28915-926),
designated HRP, and their polynucleotides (AAX09285-296). The proteins
can be produced using standard recombinant technology. The expression of
HRP is closely associated with cell proliferation and the polypeptides
and polynucleotides are useful in the diagnosis, treatment and
prevention of diseases associated with cell proliferation, particularly
immune responses and cancer. The protein or agonists may be administered
to treat or prevent a cancer such as adenocarcinoma, leukemia, lymphoma,
melanoma, myeloma, sarcoma, and tetracarcinoma. Such cancers include,
but are not limited to, cancers of the adrenal gland, bladder, bone,
brain, cervix, breast, gall bladder, ganglia, ovary and pancreas. Where
HRP is promoting leukocyte activity or proliferation, antagonists which
decrease the activity of HRP are administered. Such responses may be
associated with disorders such as Addison's disease, adult respiratory
distress syndrome, AIDS, allergies, anemia, asthma, atherosclerosis,
bronchitis, cholecystitis, Crohn's disease, ulcerative colitis, atopic
dermatitis, dermatomyositis, diabetes mellitus, emphysema, atrophic
gastritis, gout, glomerulonephritis, Grave's disease, hyperlipidemia,
irritable bowel syndrome, lupus erythematosus, multiple sclerosis,
myasthenia gravis, myocardial or pericardial infarction, osteoarthritis,

CC osteoporosis, pancreatitis, polymyositis, rheumatoid arthritis,
 CC scleroderma, Sjogren's syndrome, and autoimmune thyroiditis,
 CC complications of cancer, hemodialysis, extracorporeal circulation; viral,
 CC bacterial, fungal, parasitic, protozoan, and helminthic infections; and
 CC trauma. The HRP polynucleotide may be used for diagnosis of these
 CC conditions, and as a source of primers and probes.

XX Sequence 623 AA:

Query Match 3.9%; Score 97.5; DB 20; Length 623;
 Best Local Similarity 19.4%; Pred. No. 1.8; Mismatches 172; Indels 229; Gaps 34;

Matches 116; Conservative 82; Mismatches 172; Indels 229; Gaps 34;

DB 13 ALLFLLASWITFTVFNSTVWSALNLSLHWNSTKSLPKPTPL-----59
 DB 72 ALCGLAS-----hpnstgv-hlhlslstft-----gqllsladcklelnsgrtyglg 119
 OY 60 -----ISLKLPLTERELRIKETE--KLDQOIPRPPT-----HYNTTSATSTSTTL 105
 DB 120 nglgkmlisaikyrevpiphidylhlrtrempsdktphcymevdteramlekeerl 179
 OY 106 NPNDTCRGDQLHILEVRDHLGRKQYGSDFLRARNSPALMAGAS-----GKVTDF 158
 DB 180 ahedac--eklmelyerleel-----dackaemraarilhgjgftpmgqkklkdf 229
 OY 159 NNGTYLVSFLEWEGQVSL-----LLLIHPSEGVS--SALMSARN--OGYDPI-- 203
 DB 230 sgg-----wrmvaleralflrpfmlldpctnhldlaecvyleektkrlvl 280
 OY 204 --FTGQFVNG--TSQVHSECGILNTNMAELCOYLDNRDQ-----EGF----- 241
 DB 281 vshagdfingvcnlilnh-----nklkytgyndqgvtrtleengmkrlfwegd 333
 OY 242 -----YCVRPQH-----MPCALTHMSKNKKVSY-----266
 DB 334 qlehmkytarfghsacklarqgsketclqkmasglttervasktclsfyfpcoklpp 393
 OY 267 -----LSKQKSLEPESNNGVEIMKPFMT---SVSKCTLTK-----S 301
 DB 394 pvlmwqnvskfytkdpcplynlbfqldidervalvpgnagkscllklitgelllptcdm 453
 OY 302 VDLHESGK-----LOHQLAVDLDRNINIQMOKYCPPLIGSMYSVEMEYLRRAIDRT 354
 DB 454 lrfshvsklgyryhqlbeqldidls--pleymmkcp-----elkekeemklligy 503
 OY 355 G--GEKNIVY--ISLQGRF-----PPIDVFIKRALVHKAIOHLLLSPPDTMT 402
 DB 504 gltgkqgvspirlnsdqkrcvclawlpnpllfdepln-----hldietdal-- 555
 OY 403 IKTENIREMYNDARFSDFHGYI-----QYLIIKIDIFODLSV---SIIDAM--DITIA 451
 DB 556 -----adaInefeggmnlvshdflllqyagelwckeqcltkwpgdl-1ay 601

RESULT 14
 ID AAY31141 standard; Protein; 953 AA.
 AC AAY31141;
 DT 25-OCT-1999 (first entry)
 DE Human CARD-4L protein.

XX CARD-3; caspase recruitment domain; CARD-4; regulation; detection;
 XX caspase activation; detection; screening; therapy; diagnosis; disease;
 XX apoptotic cell death; Fas/APO-1 receptor complex; TNF receptor complex;
 XX cancer; follicular lymphoma; carcinoma; p53 mutation; viral infection;
 XX hormone-dependent tumour; autoimmune disorder; Alzheimer's disease;
 XX systemic lupus erythematosus; immune-mediated glomerulonephritis; stroke;
 XX Parkinson's disease; amyotrophic lateral sclerosis; retinitis pigmentosa;
 XX spinal muscular dystrophy; cerebellar degeneration; anaemia; drug;

KM myelodysplastic syndrome; myocardial infarction; cell proliferation;
 KM cell differentiation; cell survival; CARD-4L; CARD-4S; CARD-4V;
 KM CARD-4Z; human.

XX Homo sapiens.

XX Key Location/Qualifiers

FT 15..114 "predicted CARD domain"
 FT /note="predicted CARD domain"
 FT 197..212 "predicted kinase 1a (P-loop) domain"
 FT /note="predicted kinase 1a (P-loop) domain"
 FT 198..397 "nucleotide binding domain"
 FT /note="nucleotide binding domain"
 FT 202..209 "predicted Walker Box A"
 FT /note="predicted Walker Box A"
 FT 273..288 "predicted kinase 2 domain"
 FT /note="predicted kinase 2 domain"
 FT 280..284 "predicted Walker Box B"
 FT /note="predicted Walker Box B"
 FT 327..338 "predicted kinase 3a subdomain"
 FT /note="predicted kinase 3a subdomain"
 FT 674..950 "region containing predicted leucine-rich"
 FT /note="region containing predicted leucine-rich"
 FT 674..701 "predicted leucine-rich repeat 1"
 FT /note="predicted leucine-rich repeat 1"
 FT 702..727 "predicted leucine-rich repeat 2"
 FT /note="predicted leucine-rich repeat 2"
 FT 728..754 "predicted leucine-rich repeat 3"
 FT /note="predicted leucine-rich repeat 3"
 FT 755..782 "predicted leucine-rich repeat 4"
 FT /note="predicted leucine-rich repeat 4"
 FT 783..810 "predicted leucine-rich repeat 5"
 FT /note="predicted leucine-rich repeat 5"
 FT 811..838 "predicted leucine-rich repeat 6"
 FT /note="predicted leucine-rich repeat 6"
 FT 839..866 "predicted leucine-rich repeat 7"
 FT /note="predicted leucine-rich repeat 7"
 FT 867..894 "predicted leucine-rich repeat 8"
 FT /note="predicted leucine-rich repeat 8"
 FT 895..922 "predicted leucine-rich repeat 9"
 FT /note="predicted leucine-rich repeat 9"
 FT 923..950 "predicted leucine-rich repeat 10"
 FT /note="predicted leucine-rich repeat 10"

PN M09940102-A1.

PD 12-AUG-1999.

PF 05-FEB-1999; 99MO-US02544.

PR 08-DEC-1998; 98US-0207359.

PR 06-FEB-1998; 98US-0019942.

PR 17-JUN-1998; 98US-0099041.

(MILL-) MILLENNIUM PHARM INC.

Bertin J;

DR MPI: 1999-494269/41.

DR N-PSDB; AA209247.

XX Novel CARD-3 and CARD-4 genes and polypeptides used or treating
 XX regulation of cellular proliferation and differentiation and cell
 XX survival

PS Claim 13a; Fig 4; 181pp; English.

XX This invention describes the isolation of novel human caspase
 XX recruitment domain, CARD-3 and CARD-4 polynucleotides and proteins and a
 XX partial murine CARD-4L protein and genes. The genes and proteins of
 XX the invention are involved in the regulation of caspase activation.
 XX The caspase recruitment domain (CARD) polynucleotides, polypeptides,
 XX homologues and antibodies can be used in screening assays, detection

assays, predictive medicine and therapeutic and prophylactic methods of treatment. The methods may be used to diagnose and treat patients which are suffering from a disorder associated with abnormal level or rate of apoptotic cell death, abnormal activity of the Fas/FAO-1 receptor complex, abnormal activity of the TNF receptor complex, or abnormal activity of a caspase. Diseases that may be treated include cancer (particularly follicular lymphoma, carcinomas associated with mutations in p53 and hormone-dependent tumours), autoimmune disorders (e.g. systemic lupus erythematosus, immune-mediated glomerulonephritis), viral infections, Alzheimer's disease, Parkinson's disease, amyotrophic lateral sclerosis, retinitis pigmentosa, spinal muscular dystrophy, cerebellar degeneration, anaemia, myelodysplastic syndrome, myocardial infarction, and stroke. CARD-3 protein interacts with other cellular proteins, and so can be used for regulation of cellular proliferation and differentiation and cell survival. The CARD proteins may also be used to for screen drugs or compounds which modulate their activity. The CARD-4 gene can express a long transcript that encodes CARD-4L, a short transcript that encodes CARD-4S or two CARD-4 splice variants, CARD-4Y and CARD-4Z. This sequence represents the human CARD-4L protein described in the method of the invention.

XX Sequence 953 AA:

Query Match 3.9%; Score 97.5; DB 20; Length 953;
Best Local Similarity 20.5%; Pred. No. 3.6;
Matches 96; Conservative 58; Mismatches 154; Indels 161; Gaps 23;

```

QY 16 LASHITVFNSTKWSALNLSLHYNNSTKSLFPKTPILSKPLTELRKEITE 77
   |||||
Db 162 lfciwlltrciq-----hf-----raafegspqldpcmtldvll-vte 430
   |||||
QY 78 KLDQIPRPPTHTNTTSATHTATILNPRDYC-----RGDQHLILEVRDHL--- 127
   |||||
Db 431 vhlrmgpsalvgrrtrs-----pvelhagrdltclslgyahgmeksifvtgevgas 486
   |||||
QY 128 ---GRKKYGGGFLRARKMSPALAAGSKGVYDFNNGTYLVSTFLW---EGVSLSL 181
   |||||
Db 487 glgerdmqg--flra---lpelgpgdqgsyefhnlqlgafttaflvldrvgtqell 541
   |||||
QY 182 IHPSEGSALMSARNQY-----DRVIFGQFVNQTSQVHSE 218
   |||||
Db 542 rfqewmpagaattscypflfpqclgsgparedlfnkdhfgtnlfl----- 592
   |||||
QY 219 GGLIINTNAELCOYLDNRDQEGFCVRPHMGCAAL-----THNSKNNKVSYSKQ 270
   |||||
Db 593 cgliskakqkllrhl-----vpaalirrkakalwahllsslr--gylksl 635
   |||||
QY 271 EKSLFERSNVGVEIMKFL-----NTISVSKNTLTK----- 300
   |||||
Db 636 prvgvesfn-gvgampctflmlrclyetgsgkvglaaargicanykllycnaacsadcsa 694
   |||||
QY 301 -SYDLHESGKIQHOLAVYLDNRN-----INTQMKKCYCPPLIGSMYSKVEKEVLT 355
   |||||
Db 695 lsfvhlhfpk-----rlatldmndygyrelqpcstflvrlrlysvngl-----ldg 743
   |||||
QY 744 gvkylsecltkykyivtylgylnmqj--tdvgaryvcklldeckqthlkl 791
   |||||

```

RESULT 15
AA015552
11 AA015552 standard: Protein: 953 AA.

XX AA015552:

DT 28-FEB-2001 (first entry)

XX Apoptosis related protein encoded by gene 2 clone HDPBW68.

KW Cytostatic; antiinflammatory; immunosuppressive; antiapoptotic; cardiatic;
virtualdat; anti-AIDS; vasotrophic; anti-ischaemic; antiparkinsonian;

KW anti-Alzheimer; gene therapy; human; apoptosis; fusion protein; cancer;
KW colon; breast; prostate; melanoma; lymphoma; inflammation; herpes;
KW autoimmune disorder; multiple sclerosis; viral infection.

OS Homo sapiens.

PN MO200056752-A2.

PD 28-SEP-2000.

XX 15-MAR-2000; 2000MO-US06642.

PF 24-MAR-1999; 99US-0126018.

PR 17-JUN-1999; 99US-0139638.

PR 18-AUG-1999; 99US-0149449.

PA (HUMA-) HUMAN GENOME SCI INC.

XX Ruben SM, N1 J, Young PA;

DR WPI; 2000-587660/55.

DR N-PSDB; AAA95791.

PT Nucleic acids encoding human apoptosis associated protein, useful for
PT the prevention, treatment and diagnosis of e.g. Alzheimer's and
PT Parkinson's disease, inflammation and ischaemic injury -

XX Claim 11: Page 254-257; 273pp; English.

The invention relates to the isolation of genes encoding 9 human
apoptosis-related proteins. The nucleotide sequences AAA95790-A95798
CC encode the human apoptosis related proteins AAB1551-B1555. The genes
CC can be used to generate fusion proteins by linking to the gene for the
CC human immunoglobulin G Fc (196 Fc) portion (AAA95799) for increasing the
CC stability of the fusion protein as compared to the human protein only.
CC The gene and encoded protein may be used in the prevention, treatment
CC and diagnosis of diseases associated with inappropriate apoptosis
CC associated protein expression, e.g. cancer (e.g. colon, breast and
CC prostate cancer, melanomas and lymphomas), inflammation, autoimmune
CC disorders (e.g. multiple sclerosis) and viral infections (e.g. herpes).

XX Sequence 953 AA:

Query Match 3.9%; Score 97.5; DB 21; Length 953;
Best Local Similarity 20.5%; Pred. No. 3.6;
Matches 96; Conservative 58; Mismatches 154; Indels 161; Gaps 23;

```

QY 18 LASHITVFNSTKWSALNLSLHYNNSTKSLFPKTPILSKPLTELRKEITE 77
   |||||
Db 392 lfciwlltrciq-----hf-----raafegspqldpcmtldvll-vte 430
   |||||
QY 78 KLDQIPRPPTHTNTTSATHTATILNPRDYC-----RGDQHLILEVRDHL--- 127
   |||||
Db 431 vhlrmgpsalvgrrtrs-----pvelhagrdltclslgyahgmeksifvtgevgas 486
   |||||
QY 128 ---GRKKYGGGFLRARKMSPALAAGSKGVYDFNNGTYLVSTFLW---EGVSLSL 181
   |||||
Db 487 glgerdmqg--flra---lpelgpgdqgsyefhnlqlgafttaflvldrvgtqell 541
   |||||
QY 182 IHPSEGSALMSARNQY-----DRVIFGQFVNQTSQVHSE 218
   |||||
Db 542 rfqewmpagaattscypflfpqclgsgparedlfnkdhfgtnlfl----- 592
   |||||
QY 219 GGLIINTNAELCOYLDNRDQEGFCVRPHMGCAAL-----THNSKNNKVSYSKQ 270
   |||||
Db 593 cgliskakqkllrhl-----vpaalirrkakalwahllsslr--gylksl 635
   |||||
QY 271 EKSLFERSNVGVEIMKFL-----NTISVSKNTLTK----- 300
   |||||
Db 636 prvgvesfn-gvgampctflmlrclyetgsgkvglaaargicanykllycnaacsadcsa 694
   |||||
QY 301 -SYDLHESGKIQHOLAVYLDNRN-----INTQMKKCYCPPLIGSMYSKVEKEVLT 355
   |||||

```

Db 695 |sfv| |hfpk---r|aldldnnlnldygre|elqpcfer|elvlrlsvng|-----tdg 743
QY 356 GEX-----NTVYV|S|G|O|H|F|R|P|I|D|Y---F|I|R|R|A|N|V|H|K|A|I|O|H|L|L 394
Db 744 gvxv|seel|kyk|lvty|g|l|ynq|tdv|gar|yv|tk|l|de|ck|g|l|th|k|l 791

Search completed: October 12, 2001, 16:03:25
Job time: 48 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 12, 2001, 16:02:37 ; Search time 16.9 seconds

Title: US-09-729-454-1
Percent score: 2485
(without alignments)
962,802 Million cell updates/sec

Sequence: 1 MKTSMINYSKLLALLFLTAS.....VHPPOHVGNOINILNLYIC 475

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Statistics: 93435 swps, 74255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Query Length	DB ID	Description
1	1406	64.6	540	BB61_RABIT	P00504 oryctolagus
2	107.5	4.3	4563	APB_HUMAN	P04114 homo sapiens
3	104	4.2	694	VGLE_HSV60	P30001 human herpes
4	103	4.1	793	STL_MYCPN	P75338 mycoplasma
5	101.5	4.1	1902	P2P_LACPA	Q02470 lactobacill
6	101.5	4.1	3649	ACVS_NOCIA	P27743 nocardia la
7	100	4.0	1031	TERT_EUPAE	O00939 euploides ae
8	98.5	4.0	802	CSD2_ECOLI	P53513 escherichia
9	98	3.9	1174	NOS_RHOPE	Q26240 rhodnius pr
10	97.5	3.9	271	NPH1_RAT	O39663 ratus norv
11	97.5	3.9	623	ABR2_HUMAN	Q94963 homo sapiens
12	97	3.9	948	P1NH_ARATH	P16954 arabidopsis
13	96.5	3.9	773	GLGB_SYNP7	P52981 synecocyst
14	96	3.9	770	GLGB_SYNP3	P33539 agaricus bi
15	95.5	3.8	1102	RPOP_AGABT	P53512 escherichia
16	95	3.8	819	CSD1_ECOLI	P10978 nicotiana t
17	95	3.8	1328	POLX_TORAC	Q09111 mus musculu
18	93.5	3.8	454	GNAD_MOUSE	P52338 human herpes
19	93.5	3.8	1132	DNBI_HSV60	Q09853 schizosacch
20	93.5	3.8	1616	YABD_SCHPO	O60152 schizosacch
21	93.5	3.8	1958	UBR1_SCHPO	P16271 lactococcus
22	92.5	3.7	1902	P1P_LACLA	P33481 mumps virus
23	92	3.7	538	NPGL_MOUSE	O61200 mus musculu
24	91	3.7	253	NGFI_MOUSE	P37089 ratus norv
25	91	3.7	698	SCAA_RAT	P57504 buchnera ap
26	91	3.7	802	MUTS_BUCAI	P22538 human herpes
27	91	3.7	1132	DNBI_HSV62	Q09614 candida tto
28	90.5	3.6	627	CACP_CANTR	P47808 acanthamoeb
29	90.5	3.6	1577	MSH_ACACA	P15293 lactococcus
30	90.5	3.6	1902	P2P_LACLA	P19716 mumps virus
31	90	3.6	538	VGLE_MOUSE	O51229 borrelia bu
32	90	3.6	610	MUTL_BOBBU	

34	89.5	3.6	569	Y13_MYCE	P47385 mycoplasma
35	89.5	3.6	668	P82_YEAST	P08018 saccharomyc
36	89.5	3.6	839	NAH3_DIDNA	Q28362 dieliphis m
37	89.5	3.6	1108	CN3B_RAT	Q63085 ratus norv
38	89.5	3.6	1857	PAS2_PENFA	P15368 pentacillium
39	89	3.6	382	VMAT_SV41	P25182 simian virus
40	89	3.6	538	VGLE_MOUSE	P11236 mumps virus
41	89	3.6	699	SCAA_MOUSE	O61190 mus musculu
42	89	3.6	908	DP01_BOBBU	O51498 borrelia bu
43	89	3.6	959	VILL_DICDI	P36410 dictyostell
44	89	3.6	2787	TELI_YEAST	P8110 saccharomyc
45	88.5	3.6	230	HBP_MAIZE	P6641 zea mays (m

ALIGNMENTS

RESULT 1					
ID	BB61_RABIT	STANDARD:	PRT:	540 AA.	
AC	Q05004:				
DT	01-FEB-1994 (Rel. 28, Created)				
DT	01-FEB-1994 (Rel. 28, Last sequence update)				
DT	01-FEB-1996 (Rel. 33, Last annotation update)				
DE	BRUSH BORDER 61.9 KDA PROTEIN PRECURSOR.				
GN	ADRAH-A.				
OS	Oryctolagus cuniculus (Rabbit).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.				
OX	NCBI_TaxID=9986;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=Intestine:				
RX	MEDLINE=93286338; PubMed=8509424:				
RA	BOLL W., Schmidt-Chanda T., Semenza G., Mantel N.:				
RT	"Messenger RNAs expressed in intestine of adult but not baby rabbits.				
RT	Isolation of cognate cDNAs and characterization of a novel brush				
RT	border protein with esterase and phospholipase activity."				
RL	J. Biol. Chem. 268:12901-12911(1993).				
CC	-I- TISSUE SPECIFICITY: INTESTINE, AND IN A LESSER EXTENT IN KIDNEY.				
CC	-I- DEVELOPMENTAL STAGE: EXPRESSED IN THE INTESTINE OF ADULT BUT NOT				
CC	BB61_RABBIT.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/				
CC	or send an email to license@sib-sib.ch).				
CC	EMBL: Z12840; CA78302.1; -.				
DR	PIR: B45665; B45665.				
DR	PIR: B45665; B45665.				
KW	SIGNAL.				
FT	CHAIN 1 22				
FT	CHAIN 23 540				
FT	SEQUENCE 540 AA; 61888 MW; EB6570EBC69P25X CRC64;				

Query Match	64.6%	Score 1606:	DB 1:	Length 540:	
Best Local Similarity	57.3%	Pred No 6.3e-117:			
Matches 309;	Conservative 67;	Mismatches 91;	Indels 72;	Gaps 2;	
QY	8	YKSLALFLIASHTITFYFONSTRKWSALNISLHYNNSTKSLPKPTLISKLPLTE	67		
DB	3	HYKLLICLLALACVLCITISONSRTKMGAKLNSHYNSMTSSIPKMSVPSVSLTE	62		
QY	68	TELRIKEIIEKDOQIPPRFTHVNTTSATATILNPPOTYCRGDOLHLLLEVRHL	127		
DB	63	TELRIKEIIEKDOQLRPPRFTHTVNTTSATATILNPPOTYCRGDOLHLLLEVRHL	122		
QY	128	GRRKQYGDPLRRKSSPALMAAGSVYDFFNNGYLVSTFLFEGGVSLILLHPSEG	147		

Db 123 GHOKEYGDFLRAMFSPALKAGAGKYTFDNNGTIVSFTLFWEGVSYVLLHPSEG 182
 QY 188 VSALMSARNGCYDRVIFTCQFNGTSGVHSECLLINTNAELCOYLDNRDQEGFCVPRQ 247
 Db 183 ASALMRARNGCYDRILIFNGQFNGTSHVTECSLTLSNTECKYXLRDQDFVCMKRP 242
 QY 248 HMPGCAALTHMSKRRKVSYSLSKQESLFEPSNNGVINEKNTISVSKCN----- 297
 Db 243 HMPGCAALTHYSRRNDISYLSKSKENLFHNSKGVGEIM-KNOHIDVSOCKSKSEYKCKQ 301
 QY 298 ----- 297
 Db 302 IGKRIYVPGGYTLQGRFWLTFEQNOIQDITAKISGLKGLIYMGSTLRQMTIYLPKYM 361
 QY 298 -TLKSYDLHESGKLQHLAYLDLNNINIQOKCPPLGSGTYSVKEMEYLRALDRFTG 356
 Db 362 KTLKFPDLHETGKFKKHLLEKHTQIOKMKSHPEFTYQLFSDVIGHGIFQEDILRIG 421
 QY 357 EKNTIVISLGGHFRFPIDVIFRRALNKAIOHLLESPDVTWIKTENIREMYNDAE 416
 Db 422 DKDTVIVITPGQHFRFPIDIFIRRAISVQAIERLFRSPATKVIYKTEINIREMHTAE 481
 QY 417 RPSDFHGYIOYLIDIFODISLIDADITATGTTNNVHPQHYVGNQILLNTYC 475
 Db 482 RFGDFHGYIOYLIDIFKDLNKGVDAMDITATGTTNNVHPQHYVGNQILLNTYC 540
 RESULT 2
 APB_HUMAN STANDARD: PRT: 4563 AA.
 ID APB_HUMAN 000502: Q13787;
 AC P04114: 000502: Q13787;
 DT 01-NOV-1986 (Rel. 03, Created)
 DT 01-NOV-1986 (Rel. 03, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE APOLIPOPROTEIN B-100 PRECURSOR (APO B-100) [CONTAINS: APOLIPOPROTEIN B-48 (APO B-48)].
 GN APOB.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_Taxid:9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-87016385; PubMed-3763409;
 RA Knott T.C., Wallis S.C., Powell L.M., Pease R.J., Lusis A.J., Blackhart B., McCarthy B.J., Mahley R.W., Levy-Wilson B., Scott J.;
 RT "Complete cDNA and derived protein sequence of human apolipoprotein B-100.";
 RL Nucleic Acids Res. 14:7501-7503(1986).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-88003974; PubMed-3652907;
 RA Luoway E.H., Blackhart B.D., Pierotti V.R., Calati L., Fortier C., Knott T., Scott J., Mahley R.W., Levy-Wilson B., McCarthy B.J.;
 RT "DNA sequence of the human apolipoprotein B gene.";
 RL DNA 6:363-372(1987).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-87008488; PubMed-375943;
 RA Chen S.-H., Yang C.-Y., Chen P.-F., Setzer D., Tanimura M., Li W.-H., Goto A.M., Jr., Chan L.;
 RT "The complete cDNA and amino acid sequence of human apolipoprotein B-100.";
 RL J. Biol. Chem. 261:12918-12921(1986).
 RN [4]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-87041416; PubMed-3464946;
 RA Law S.W., Grant S.M., Higuchi K., Hospattankar A.V., Lackner K.J., Lee N., Brewer H.B., Jr.;
 RT "Human liver apolipoprotein B-100 cDNA: complete nucleic acid and derived amino acid sequence.";
 RL Proc. Natl. Acad. Sci. U.S.A. 83:8142-8146(1986).
 RN [5]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-87161758; PubMed-3030729;
 RA Cladas C., Hadzopoulou-Cladas M., Nolte R.T., Atkinson D., Zannis I.;
 RT "The complete sequence and structural analysis of human apolipoprotein B-100: relationship between apoB-100 and apoB-48 forms.";
 RL EMBO J. 5:3495-3507(1986).
 RN [6]
 RP SEQUENCE OF 709-906 FROM N.A.
 RX MEDLINE-85270450; PubMed-3860836;
 RA Deed S.S., Motulsky A.G., Alberts J.J.;
 RT "A partial cDNA clone for human apolipoprotein B.";
 RL Proc. Natl. Acad. Sci. U.S.A. 82:4983-4986(1985).
 RN [7]
 RP SEQUENCE OF 3056-3159 FROM N.A.
 RX MEDLINE-86041888; PubMed-3903660;
 RA Mehrabian M., Schumaker V.N., Fareed G.C., West B., Kirchgessner T.G., Lin H.-C., Wang X., Ma Y., Mendez E., Lusis A.J.;
 RT "Human apolipoprotein B: identification of cDNA clones and characterization of mRNA.";
 RL Nucleic Acids Res. 13:6937-6953(1985).
 RN [8]
 RP SEQUENCE OF 1937-2018 AND 3811-4334 FROM N.A.
 RX MEDLINE-86093680; PubMed-3841204;
 RA Carlsson P., Olsson S.O., Bondiers G., Darnet S., Wiklund Burell G.;
 RT "Molecular cloning of human apolipoprotein B cDNA.";
 RL Nucleic Acids Res. 13:8813-8826(1985).
 RN [9]
 RP SEQUENCE OF 3109-4563 FROM N.A.
 RX MEDLINE-85300528; PubMed-2944225;
 RA Knott T.J., Rall S.C., Jr., Innerarity T.L., Jacobsen S.F., Udeas M.S., Levy-Wilson B., Powell L.M., Pease R.J., Eddy R., Nakai A., Byers M., Priestley L.H., Robertson E., Pali L.H., Betsholtz C., Shows T.B., Mahley R.W., Scott J.;
 RT "Human apolipoprotein B: structure of carboxyl-terminal domains sites of gene expression, and chromosomal localization.";
 RL Science 230:37-43(1985).
 RN [10]
 RP SEQUENCE OF 1-291 FROM N.A.
 RX MEDLINE-86149325; PubMed-3513177;
 RA Proter A.A., Hardman D.A., Schilling J.W., Miller M., Proter A.A., Hardman D.A., Schilling J.W., Miller M.;
 RT "Isolation of a cDNA clone encoding the amino-terminal region of human apolipoprotein B.";
 RL Proc. Natl. Acad. Sci. U.S.A. 83:1467-1471(1986).
 RN [11]
 RP SEQUENCE OF 1-1670 FROM N.A.
 RX MEDLINE-86287319; PubMed-3461454;
 RA Proter A.A., Hardman D.A., Sato K.Y., Schilling J.W., Miller M., Yamanaka M., Hori Y.J., Hjerlid K.A., Chen G.C., Kane J.P.;
 RT "Analysis of cDNA clones encoding the entire hApoB-100 protein.";
 RL Proc. Natl. Acad. Sci. U.S.A. 83:5678-5682(1986).
 RN [12]
 RP PARTIAL SEQUENCE, AND IDENTIFICATION (APO-B48).
 RX MEDLINE-88018019; PubMed-3659919;
 RA Chen S.-H., Habib G., Yang C.-H., Gu Z.-W., Lee B.-H., Wang S.-H., Silbermann S.R., Cai S.-J., Deshpande J.F., Rosenbaum M., Goto A.M., Jr., Li W.-H., Chan L.;
 RT "Apolipoprotein B-48 is the product of a messenger RNA with an open specific in-frame stop codon.";
 RL Science 238:363-366(1987).
 RN [13]
 RP DOMAINS.
 RX MEDLINE-87039351; PubMed-3773997;
 RA Knott T.C., Pease R.J., Powell L.M., Wallis S.C., Rall S.C., Jr., Innerarity T.L., Blackhart B., Taylor W.R., Marcei F., Milne R., Johnson D., Fuller M., Wuehls A.J., McCarthy B.J., Mahley R.W., Levy-Wilson B., Scott J.;
 RT "Complete protein sequence and identification of structural domains of human apolipoprotein B."

RI Nature 323:734-738(1986).
 RP [14]
 RP DUMAINS.
 RA Y and C.-Y., Chen S.-H., Giansurco S.H., Bradley W.A., Sparrow J.T.,
 RA Tanimura M., Li W.-H., Sparrow D.A., Deloof H., Rosseneu M.,
 RA Lee F.-S., Gu Z.-W., Gotto A.M. Jr., Chan L.;
 RI "Sequence, structure, receptor-binding domains and internal repeats
 RI of human apolipoprotein B-100.";
 RI Nature 323:738-742(1986).
 RI [15]
 RP CALCIUM-BINDING DATA.
 RX MEDLINE-86242245; PubMed-3087360;
 RA Dashi N., Lee D.M., Mok T.;
 RI "Apolipoprotein B is a calcium binding protein.";
 RI Biochem. Biophys. Res. Commun. 137:493-499(1986).
 RI [16]
 RP VARIANT SER-4338.
 RX MEDLINE-91071750; PubMed-1979313;
 RA Navajas M., Laurent A.-M., Koreel J.-F., Ragab A., Cambou J.-P.,
 RA Cuny G., Cambien F., Rolzes G.;
 RI "Detection by denaturing gradient gel electrophoresis of a new
 RI polymorphism in the apolipoprotein B gene.";
 RI Hum. Genet. 86:91-93(1990).
 RI [17]
 RP VARIANT FDB GLN-3527.
 RX MEDLINE-89098975; PubMed-2563166;
 RA Sofia L.F., Ludwig E.H., Clarke H.R.G., Vega G.L., Grundy S.M.,
 RA McCarthy B.J.;
 RI "Association between a specific apolipoprotein B mutation and
 RI familial defective apolipoprotein B-100.";
 RI Proc. Natl. Acad. Sci. U.S.A. 86:587-591(1989).
 RI [18]
 RP VARIANT LEU-2739.
 RX MEDLINE-91016974; PubMed-2216805;
 RA Huang L.-S., Gavish D., Breslow J.L.;
 RI "Sequence polymorphism in the human apoB gene at position 8344.";
 RI Nucleic Acids Res. 18:5922-5922(1990).
 RI [19]
 RP VARIANT FDB CYS-3558.
 RX MEDLINE-95190020; PubMed-7883971;
 RA Pullinger G.R., Hennessey L.R., Chatterton J.F., Liu W., Love J.A.,
 RA Mendel G.M., Frost P.H., Malloy M.J., Schumaker V.N., Kane J.P.;
 RI "Familial ligand-defective apolipoprotein B. Identification of a new
 RI mutation that decreases LDL receptor binding affinity.";
 RI J. Clin. Invest. 95:1225-1234(1995).
 RI [20]
 RP VARIANTS L-1437; S-1914; K-2566; T-3121; A-3945; M-4128 AND T-4481.
 RX MEDLINE-97044521; PubMed-8889592;
 RA Poirier O., Ricard S., Behague I., Souriau C., Evans A.,
 RA Atwell D., Marques-Vidal P., Luc G., Rolzes G., Cambien F.;
 RI "Detection of new variants in the apolipoprotein B (Apo B) gene by
 RI PCR-SSCP.";
 RI Hum. Mutat. 8:282-285(1996).
 RI [21]
 RP VARIANTS FDB GLN-3527 AND CYS-3558.
 RX MEDLINE-97403938; PubMed-9259199;
 RA Rabes J.P., Vaeret M., Saint-Jore B., Erlich D., Jondeau G.,
 RA Krempf M., Giraudet P., Junien C., Boileau C.;
 RI "Familial ligand-defective apolipoprotein B-100: simultaneous
 RI detection of the ARG3500->Gln and ARG3531->Cys mutations in a
 RI French population.";
 RI Hum. Mutat. 10:160-161(1997).
 RI [22]
 RP VARIANTS S-1914; R-1923; L-2739; D-3319; T-3427; Q-3432 AND I-3921.
 RX MEDLINE-98141125; PubMed-9490296;
 RA Lefron T.P., Bakken K.S., Hoel V., Hjermann I., Berg K.;
 RI "Screening for mutations of the apolipoprotein B gene causing
 RI hypcholesterolemia.";
 RI Hum. Genet. 102:44-49(1998).
 RI [23]
 RP FUNCTION: APOLOPROTEIN B IS A MAJOR PROTEIN CONSTITUENT OF
 RP CHYLOMICRONS, VLDL AND LDL. IT FUNCTIONS AS A RECOGNITION SIGNAL
 RP FOR THE CELLULAR BINDING AND INTERNALIZATION OF LDL PARTICLES BY
 RP THE APOB/E RECEPTOR.

CC -1- DISEASE: DEFECTS IN APOB ARE A CAUSE OF FAMILIAL LIGAND-DEFECTIVE
 CC APOLOPROTEIN B-100 (FDB). IT IS A DOMINANTLY INHERITED DISORDER
 CC OF LIPOPROTEIN METABOLISM LEADING TO HYPERCHOLESTEROLEMIA AND
 CC INCREASED PRONESS TO CORONARY ARTERY DISEASE (CAD).
 CC THE PLASMA CHOLESTEROL LEVELS ARE DRAMATICALLY ELEVATED DUE TO
 CC IMPAIRED CLEARANCE OF LDL PARTICLES BY DEFECTIVE APOB/E RECEPTORS.
 CC -1- DISEASE: DEFECTS IN APOB ASSOCIATED WITH OTHER GENE DEFECTS
 CC (POLYGENIC) CAN BE THE CAUSE OF HYPOCHOLESTEROLEMIA.
 CC MISCELLANEOUS: APO B-48 WHICH IS SYNTHESIZED ONLY BY THE
 CC INTESTINE, AND FOUND IN CHYLOMICRONS, IS A SHORTENED FORM OF APO
 Query Match 4.38; Score 107.5; DB 1; Length 4563;
 Best Local Similarity 20.78; Pred. No. 14;
 Matches 90; Conservative 71; Mismatches 169; Indels 105; Gaps 22;
 QY 56 KTLPLSLKPLTE-LRIKETIEKLDQIFPPPTFH-----NTTSATHS-TATILNPRD 109
 DB 2025 KVPLLSEPIITIDALEMRDAVER-----PQETIVAFYVKDKNDVSNLPPFTIQ 2078
 QY 110 TYCGDQLHLEVDHGRKQYGD-FLRRKSSPALMAAGSKVTFDNGTYLVST 168
 DB 2079 EYFRNRQTIIVVENVORNLKHNIDQFVRK-----YRALGKLPQAN-DYLSNFN 2130
 QY 169 LMEGQVSLLLIHPSEGVSAIWS-----YRAGGDRVIF 204
 DB 2131 --MERQVS-----HAKKFLALTKKRYRTENDIDALDDAKINFKEKLSQLOTY--MIO 2180
 QY 205 TCGPVNTQVHSECGILTNMELQYIDNDQDGRYCYRPOHMCALTRHYSK---- 260
 DB 2181 FDQYINDSDLDLKLAIANIDITIEKKSIDEN--YHRAVLVYTIHDLLTFENIDP 2238
 QY 261 NKVSYLSKQESLFEKSNVGEVMEKFTISVSCNTLKVSL-HESGKL-QHQLAVDL 318
 DB 2239 NKSQSSYASMIQNDVTKYQIRIQIKQLQULK----RHINIDIQHLAGKLNKHEIDV 2294
 QY 319 DRNINIQMKCYRPLISMTYSKMEHYLAIIDRTGKENYIVYSLAQHFRPPY--- 375
 DB 2295 -----KVLDDLQI-----TISFET-----NDVL-----EYKHFVNL 2325
 QY 376 -DVFIRALNVKAIQHLRLSPD-----MYITKTENIREMYNDAEKRSFHCYIYLT 429
 DB 2326 GDEYVAKKINAFRAKVELIEYEVDOQIYVLMKVELNHQYKKEITQKLSNVLDQV 2385
 QY 430 IKDIFODLSVSTIQA 444
 DB 2386 IKDYFEKLVGFIIDA 2400
 RESULT 3
 VGLH_HSV6U STANDARD; PRT: 694 AA.
 AC P30001;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DE 01-OCT-2000 (Rel. 40, Last annotation update)
 DE GLYCOPROTEIN B PRECURSOR.
 GN GH OR U48 OR LF2 OR BHLFI.
 OS Human herpesvirus (type 6 / strain Uganda-1102) (HHV6), and
 OS human herpesvirus (type 6 / strain GS) (HHV6).
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 OC Betaherpesvirinae; Roseolovirus.
 OC NCBI_TaxID=10370, 10369;
 RP [1]
 RP SEQUENCE FROM N.A.
 RP STRAIN=UGANDA-1102;
 RC STRAIN=UGANDA-1102;
 RX MEDLINE-93091236; PubMed-133836;
 RA Gompels U.A., Cars A.L., Sun N., Arrand J.R.;
 RI "Infectivity determinants encoded in a conserved gene block of human
 RI herpesvirus-6.";
 RI J. Virol. 66:103-109(1992).
 RI [2]
 RP DNA Seq. 3:25-39(1992).
 RP SEQUENCE FROM N.A.
 RP STRAIN=UGANDA-1102;


```

Db 902 MORTLLNINISVQVLLDGGKNTLLSSSTNQTKTYDAHS----- 943
QY 325 OMOKYCY---PL-----IGSMYSVKMEYLTRAIDRTGGEKNTYVI 364
Db 944 ---OKIYYNAPAMDGTYTDGRCNIXRADGOSTYTIASVTP-----EGGKRGVDFV 993
QY 365 SIGGHRFPIDVFIRALNWKAIQHLILRSPDTNVIKTEINR-EMYNDAERSDFG 423
Db 994 -----PFKLD-----SKAPYRIVHALSA-----KTENCKTGYLTAEADODLSG 1032
QY 424 YIQYLIKIDIFOD---LSVSIIDAMDTIYAGTNVHPP-----QHYVGNQIN 468
Db 1033 LDATSKVTAINEVTNLDTATFDAG--TTADGTYKIEPLDSDQALGNGDN 1083

RESULT 6
ACVS_NOCOA STANDARD: PRT: 3649 AA.
ID ACVS_NOCOA
AC P27743;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE DELTA-(L-ALPHA-AMINODIPYL)-L-CYSTEINYL-D-VALINE SYNTHETASE
DE (EC 6.3.1.1) (ACV SYNTHETASE) (ACVS).
GN PCBAB
OS Nocardi lactamdurans.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae.
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1913;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-VAR LC 411;
RA MEDLINE=92065808; PubMed=1956290;
RT Coque J.J.R., Martin J.F., Calzada J.G., Hiras P.;
RT "The cepamycin biosynthetic genes pcbaa, encoding a large
RT multidomain peptide synthetase, and pcob of Nocardi lactamdurans are
RT clustered together in an organization different from the same genes
RT in Acromonium chrysogenum and Penicillium chrysogenum.";
RL Mol. Microbiol. 5:1125-1133(1991).
CC -!- FUNCTION: EACH OF THE CONSTITUTED AMINO ACIDS OF THE TRIPEPTIDE
CC ACV ARE ACTIVATED AS AMINOCY-ADENILATES WITH PEPTIDE BONDS
CC FORMED THROUGH THE PARTICIPATION OF AMINO ACID THIOLESTER
CC INTERMEDIATES.
CC -!- COFACTOR: CONTAINS 3 COVALENTLY BOUND PHOSPHOPANTETHEINES.
CC -!- PATHWAY: FIRST STEP IN THE BIOSYNTHESIS OF PENICILLIN AND
CC CEPHALOSPORIN.
CC -!- SIMILARITY: TO OTHER ENZYMES WHICH ACT VIA AN ATP-DEPENDENT
CC COVALENT BINDING OF AMP TO THEIR SUBSTRATE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcements/
CC or send an email to license@ebi.ac.uk).
CC -----
DR EMBL: X57310; CAA40561.1; -
DR PIR: S18268; S18268.
DR HSP: P14687; IAKU.
DR InterPro: IPR000255; -
DR InterPro: IPR000873; -
DR InterPro: IPR001242; -
DR Pfam: PF00501; AMP-binding; 3.
DR Pfam: PF00668; DUF; 3.
DR Pfam: PF00975; Thioesterase; 1.
DR Pfam: PF00550; PP-binding; 3.
DR PRINTS: PR00134; AMP-binding; 3.
DR PROSITE: PS00012; PHOSPHOPANTETHEINE; 2.
DR PROSITE: PS00435; AMP-BINDING; 1.
DR PROSITE: PS0075; ACP_DOMAIN; 3.
KW ligase; Antibiotic biosynthesis; Multifunctional enzyme;

```

```

KW Repeat; Phosphopantetheine.
FT REPEAT 401 861
FT REPEAT 1014 1937
FT REPEAT 2079 2985
FT DOMAIN 788 857
FT DOMAIN 1864 1933
FT DOMAIN 2910 3261
FT BINDING 820 820
FT BINDING 1896 1896
FT BINDING 2944 2944
FT ACT_SITE 3502
SQ SEQUENCE 3649 AA; 404079 MM; 66D045704F856B6B C0604;

Query Match
Best Local Similarity 4.18; Score 101.5; DB 1; Length 3649;
Matches 85; Conservative 54; Mismatches 141; Indels 14; Gaps 1.

QY 102 ATILNPADY--CRGDHLILLEVRHLCRRKQYGGDFLARRHSSPALMAG-ASGKTYP 158
Db 1083 AGLLNKSRYRVSGLDYDH---YORHTRKILYLADL-----TALKAGCALQVT- 1180
QY 159 NNGTYLSTFL-MEGVSLSLILHPSEG-----VSALMSANNQGYDRIYTVGVNGI 212
Db 1131 -----LHSLVLOFVWH-----KYLAIIGGNTIYGVYSGHNLVD----- 1165
QY 213 SVHSEGGILINTNAELCOYLDRNDEGEYCVFQHMPCALITHYSKKKVSLSKER 27
Db 1167 -GIENSAGLFINI---LPLIVDHQAGGNVAENRDLQAAVNTMSKS--IVELGLUS 1220
QY 273 SFERSNVGVELMEKFNISVSKNTLKSVDLHESGKLCYUAVLDRININLQKQYGP 31
Db 1221 GEKKR-----RLFDLVLVE-NIPLILD--EEELAQEALNFER----- 1277
QY 333 LISSMTYSKMEYLTRAIDRTGGEKNTYVLSGHRFPPIDF--IRNL----- 364
Db 1258 -----AYADKVDYPIAVVARREGDELTYLTVAGELFDDIDITLLVANTLKQVOTD 1312
QY 384 -----VWKAIOHLLSPDTNVIKTEINR--- 407
Db 1313 IANPVELDILSPMKARFDSMNWETKEEPPADKTLHAFVEKMERPEIIVYKRNKL 137
QY 408 IRMYNDAREFSDPHGYIOTL-----IINDIFODLSVSIILAMDTIAY 451
Db 1373 YRELNERANKRLAHYLRSVVELRPDQVALVLRKSELMTIATIAAKTAAV 1427

RESULT 7
TERT_EUPAE STANDARD: PRT: 1031 AA.
ID TERT_EUPAE
AC 000939;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE TELOMERASE REVERSE TRANSCRIPTASE (EC 2.7.7.7) (TEL-MERASE CATALYTIC
DE SUBUNIT) (TELOMERASE SUBUNIT P122).
OS Euploates aediculatus.
OC Eukaryota; Alveolata; Ciliophora; hypotrichs; Euplotidae; Euplotidae;
OC Euploates.
OX NCBI_TaxID=5940;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=97274210; PubMed=9110970;
RA Linder J., Hughes T.R., Shevchenko A., Mann M., Lundblad V.,
RA Cech T.R.;
RT "Reverse transcriptase motifs in the catalytic subunit of
RT telomerase."
RL Science 276:561-567(1997).
CC -!- FUNCTION: TELOMERASE IS A RIBONUCLEOPROTEIN ENZYME ESSENTIAL FOR
CC THE REPLICATION OF CHROMOSOME TERMINI IN MOST EUKARYOTES. IT
CC ENCATONATES TELOMERES. IT IS A REVERSE TRANSCRIPTASE THAT AIDS
CC SIMPLE SEQUENCE REPEATS TO CHROMOSOME ENDS BY COPYING A TEMPLATE
CC SEQUENCE WITHIN THE RNA COMPONENT OF THE ENZYME.

```

```

CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- SIMILARITY: BELONGS TO THE REVERSE TRANSCRIPTASE FAMILY.
CC TELOMERASE SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: 095964; AAC47515.1;
CC Transferase: RNA-directed DNA polymerase; Telomerase; Nuclear protein;
CC DNA-binding.
CC SEQUENCE 1031 AA; 122562 MW; 57B87A63A1EED60F CRC64;

Query Match
Best Local Similarity 4.0%; Score 100; DB 1; Length 1031;
Matches 106; Conservative 71; Mismatches 179; Indels 136; Gaps 29;

UY 18 LASWILFTFQNSTKYWSALNLSISLHWNSTKSLPEKTPPLSL-KPLETELRI---K 73
UY 30 LYSWJ-----QKYLRCNOSOS-HYKLEDEIKIPAGINIVATPRDYNEEDKVIARK 80
UY 74 EI-----IEKLDQIIPRPFTHTVNTTSATHTATILNPRTDYGQDL---HILLEVR 124
UY 81 KVFSTGLMIELIDKCL-----VELSSSDVSROKLOCFGFLQKGLAKTKHLLTA-- 131
UY 125 DILGRRKUYGGDPLRAHSSPALMAGASGVKVFENNQTYLVSEFLFMGQVSLILLHP 184
UY 132 --LSTQKUY--PFODEMNOVRAMIG-----NELPEHLYTKY-----LIFORT 169
UY 185 SEGVSAALMSANOGDRAVIFGQF---VNGTSOVH-----SEGLILNTAALCYOLDN 235
UY 170 SEGLTVQPCGNV-FDHLKYNDRKFKKOKGGAAMNPPRCSTCK--YVNKNEHDHFLNK 226
UY 216 KDQEG-----FYCVAPQHPMCAALTHMYSKKKVSYLSKQESLPSRSNVGEIM 285
UY 227 INVPMNNMKSKPIFYC-----THF--RNNOQPKKHE--EVSNNKNISAM 269
UY 286 EKFNLT--SVSKCNTL-KSVVDHESGKLOHQLAVDDRNININOMQKCYCPILGSMYSAK 342
UY 270 IRAQITITFINIFRRRIKPKLKDVKYETKIAVLEKVFQNFNITYLKSC-PLPEKMERKO 328
UY 343 EMEYLTALIDTSGEK-----NTVIYISLQGHF-RPPIDVETIR----- 381
UY 329 KIEFNL--INKTREKSKYEELEFSYTDNKCVTQFINFEFYNILPKDFLGRNRKRFOK 385
UY 382 -----ALNVHKAFOHLLRSPDTMV-----IKTENIREMNDARESDPHGYIYLI 430
UY 386 KYKKYVELNKKHLLHKNLLEKINTRELISMOVETSAKHFFYFDHENI-----YVLKLL 440
UY 431 KDIFDLVSSTI 442
UY 441 RWIFEDLVSLI 452

RESULT 8
CS22_ECOLI STANDARD: PRT: 802 AA.
AC P53513.
UT 01-OCT-1996 (Rel. 34, Created)
UT 01-OCT-1996 (Rel. 34, Last sequence update)
UT 01-OCT-1996 (Rel. 34, Last annotation update)
DE OUTER MEMBRANE USHER PROTEIN CSSD PRECURSOR (CS6 FIMBRIA USEH
DE PROTEIN).
DE
DE CS22.
DE Escherichia coli.
CC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
CC Escherichia.
CC NCB1_TaxID=562;

```

```

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=88775;
RA Wolf M.K., de Haan L.A.M., Casseels F.C., Willshaw G.A.,
RA Gestel E.C.M., Gaastra W., Warren R., Boedecker E.C.;
RL Submitted (Jan-1994) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: INVOLVED IN THE EXPORT AND ASSEMBLY OF C6 FIMBRIAL
CC SUBUNIT ACROSS THE OUTER MEMBRANE.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. OUTER MEMBRANE
CC (BY SIMILARITY).
CC
CC -1- SIMILARITY: BELONGS TO THE FIMBRIAL EXPORT USHER FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: 004846; AAB51364.1;
CC InterPro: IPR000015;
CC Pfam: PF00577; Usher; 1.
CC PROSITE: PS01151; FIMBRIAL_USHER; FALSE NEG.
CC Outer membrane; Transmembrane; Fimbria; Transport; Signal.
CC SIGNAL 1
CC CHAIN 1 802 ? OUTER MEMBRANE USHER PROTEIN CSSD.
CC SEQUENCE 802 AA; 90393 MW; 5BF0DB5F1F5B5B0 CRC64;

Query Match
Best Local Similarity 4.0%; Score 98.5; DB 1; Length 802;
Matches 108; Conservative 71; Mismatches 162; Indels 205; Gaps 26;

UY 20 SWILFTVQ-----NSTKW-----SALSLSLHWNNS----- 49
UY 50 -----TKSLFPKPLISLPLETELRIKE--IEKLDQIIPRPFTHTVNTTSATHS 100
UY 249 TQDWTYTPSELDITPELHGVADSDQIIVRQGNITII--INESVPAQGFSPITNLMYT-- 304
UY 101 TATILNPRTDYGQDLHILEVRDHLGRRKQYGGDFLRARSSPALMAGA-----S 152
UY 305 -----GGQLNV--EITDIYGNKKQY-----TVNNSSLPVMRAGLMYNFTS 344
UY 153 GAVT-----DENNGTYLVSEFLFMGQVSLSL-----LIH 183
UY 345 GKLTAKKNEQSDDFGTQGDINTGYTHNS-TLFGGYQSKYFNLSTGIGTDLGSGAWLH 403
UY 184 PSE-----GVSAALMSANOGDRAVIFGQFVNGTSOVHSEC 219
UY 404 VRSNFKKKNQYINILQONTLPRLPNAGVDFDAVYKRRKYELSDIMGHNLNOLKNSF 463
UY 220 GLILNTNA-----ELCOYLDNRQEGFPCVAP--QHPPCAALTHMYSKKKVSYL 267
UY 464 SLLSKSLNKGKNSLDVYKMKMYQMDVANSNSIVFPFKFRR-AMITNCSLNKYQSYE 522
UY 268 SKQESLPSRSNVGEIMEKFTY-----SVSKCNT--LKSVDHESGKLOHQLAVDDR 320
UY 523 KDKRFFSI--NISLPLTDYGHISNYSFNSNNTATATSSVGLNGS-----FFNDARL 573
UY 321 NINIQOMK-----YCPY---LIGSMYVMEYELTALIDRTGKEKT 360
UY 574 MNNIQONNTRNNGYTDNTSYIATSYASPGVFTSSYSKMK-----YSSQYVSSAGS 626
UY 361 VIVISLQGHPRPPIDVIFIRALNVHKAIOHLLRSPDTMYITENIRMT--NDAERF 418
UY 627 IYLSHSDVAFTQ-----KAGDTSLAVRIDISIKIGNTPGVY 664
UY 419 SDFHGY 424
UY 665 TGYNGF 670

```

```

RESULT 9
NOS_RHOPR
ID NOS_RHOPR STANDARD: PRT: 1174 AA.
AC 026240;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DE NITRIC-OXIDE SYNTHASE, SALIVARY GLAND (EC 1.14.13.39) (NOS).
OS Rhodnius prolixus.
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Palaeoptera; Hemiptera; Euhemiptera;
OC Heteroptera; Pantheroptera; Cimicomorpha; Reduviidae; Triatominae;
OC Rhodnius.
OK NCBI_TaxID=13249;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Salivary gland.
RX MEDLINE=97175053; PubMed=9022713;
RA Yuda M., Hirai M., Mura K., Matsumura H., Ando K., Chanzel Y.;
RT *cDNA cloning, expression and characterization of nitric-oxide
RT synthase from the salivary glands of the blood-sucking insect
RT Rhodnius prolixus.
RL Eur. J. Biochem. 242:807-812(1996).
CC -1- FUNCTION: PRODUCES NITRIC OXIDE (NO) WHICH IS A MESSENGER MOLECULE
CC WITH DIVERSE FUNCTIONS THROUGHOUT THE BODY. THE PRODUCTION OF NO
CC IN THE SALIVARY GLAND IS USED AS A VASODILATOR FOR BLOOD SUCKING.
CC -1- CATALYTIC ACTIVITY: L-ARGININE + N NADPH + M O(2) - CITRULLINE +
CC NITRIC OXIDE + N NADP(+).
CC -1- COFACTOR: HEME. BINDS ONE MOLE EACH OF FAD AND FMN (BY
CC SIMILARITY).
CC -1- ENZYME REGULATION: STIMULATED BY CALCIUM/CALMODULIN.
CC -1- SIMILARITY: BELONGS TO THE NOS FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U59389; AAB03810.1; -
DR HSSP: P29477; 2NOS.
DR Interpro: IPR001094; -
DR Interpro: IPR001433; -
DR Interpro: IPR003097; -
DR Pfam: PF00667; FAD_binding_1.
DR Pfam: PF00175; oxidored_1ad; 1.
DR PRINTS: PR00369; FLAVODOXIN.
KW Oxidoreductase; NADP; FAD; FMN; Calmodulin-binding; Heme.
FT DOMAIN 35 45 POLY-GLN.
FT BINDING 162 162 HEME (BY SIMILARITY).
FT DOMAIN 475 495 CALMODULIN-BINDING (POTENTIAL).
FT NP_BIND 639 670 FMN (PYRIMIDINE PART) (BY SIMILARITY).
FT NP_BIND 780 791 FAD (ADP PART) (BY SIMILARITY).
FT NP_BIND 923 933 FAD (FLAVIN PART) (BY SIMILARITY).
FT NP_BIND 998 1016 NADP (RIBOSE PART) (BY SIMILARITY).
FT NP_BIND 1095 1110 NADP (ADP PART) (BY SIMILARITY).
SO SEQUENCE 1174 AA; 132393 MW; C32FE64EE51409CF CRC64;

Query Match
Best Local Similarity 3.9%; Score 98; DB 1; Length 1174;
Matches 51; Conservative 53; Mismatches 77; Indels 82; Gaps 13;
OY 208 FVNGTSOVHSECGILNLTAELCOYLDRDQEGYCVRPQHPCALALHYSKNNKAVSL 267
DB 708 FLEATOMLHSEA-----VTASTVRFVESATOD-----LCAALSHLHK----- 745
OY 268 SKQSKLFRSNVGVEMKFTNISVSKNTLSVDSKGLDHOALAVALDNRINIQMO 327

```

```

RESULT 10
NPHL_RAT
ID NPHL_RAT STANDARD: PRT: 271 AA.
AC 063366;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE NEUREXOPHILIN 1 PRECURSOR (NEUROPHILIN).
GN NXP1L OR NPH1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OK NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=96285495; PubMed=8699246;
RA Petrenko A.G., Ulrich B., Missler M., Krashinsky V., Rosahl J.W.,
RA Suedhof T.C.;
RT "Structure and evolution of neurexophilin."
RT J. Neurosci. 16:4360-4369(1996).
CC -1- FUNCTION: MAY BE SIGNALING MOLECULES THAT RESEMBLE NEURECEPTORS
CC AND THAT ACT BY BINDING TO ALPHA-NEUREXINS AND POSSIBLY OTHER
CC RECEPTORS (POTENTIAL).
CC -1- SUBCELLULAR LOCATION: SECRETED (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE NEUREXOPHILIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L27867; AAB18420.1; -
DR Glycoprotein; signal.
KW SIGNAL
FT CHAIN 1 271 POTENTIAL.
FT CARBOHYD 23 23 NEUREXOPHILIN 1.
FT CARBOHYD 68 68 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 93 93 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 146 146 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 156 156 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 162 162 N-LINKED (GLCNAC. . .) (POTENTIAL).
SO SEQUENCE 271 AA; 31017 MW; 472D12D0B071E97A CRC64;

Query Match
Best Local Similarity 3.9%; Score 97.5; DB 1; Length 271;
Matches 46; Conservative 25; Mismatches 72; Indels 51; Gaps 7;
OY 19 ASWITFYFONSTRKWSALN-----SISLHWNSTRKSLFKTPILSLK 64
DB 4 ACWYVLLDLPYVLTATCNLTNGKSKSLKSTLKHWTSSKDL-----SIS 56
OY 64 PLTELRKEIKEDDOOT-PPRPTFHVTTTSATSHATILMPDTVCNLTJLHLE 122

```

```

106 57 FULLSTFKENGITDLDLRYDTPEPPYSDOLMDLNRSTD-----LOE 99
QY 123 VRDHLGR-----KOYG-GDPLRARKMSPALMAGSGVYDFPNNGIYLVSTLFE 172
107 100 PPRAKRPPYKTKGKKRKGDF-HSNIKYKLMELITGKIVDHGIFSVYFRHNS 158
QY 173 GGVSLLELLHPSE 186
108 159 GGVNSVSLVPTK 172
RESULT 11
ABR2_HUMAN
ID ABR2_HUMAN STANDARD: PRT; 623 AA.
AC U90563; 060864;
DT 01-OCT-2000 (Rel. 40, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE AIP-BINDING CASSETTE, SUB-FAMILY F, MEMBER 2 (IRON INHIBITED ABC
DE TRANSPORTER 2).
GN ABCF2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN 11
RP SEQUENCE FROM N.A.
KC TISSUE=Uterus;
RA Ansoez W., Winkler U., Mewes H.-W., Gassenhuber J., Wilmann S.;
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
KN 12
KP SEQUENCE FROM N.A.
RX MEDLINE=20403900; PubMed=10944468;
RA Ye Z., Connor J.R.;
RT *cDNA cloning by amplification of circularized first strand cDNAs
RT reveals non-IRE-regulated iron-responsive mRNAs.*
RL Biochem. Biophys. Res. Commun. 273:223-227(2000).
RN 13
KP SEQUENCE OF 74-623 FROM N.A.
KC TISSUE=Lung;
RX PubMed=11124703;
RA Stanchi F., Bertocco E., Toppo S., Dioguardi R., Simionati B.,
RA Cannata N., Zimbelli R., Lanfranchi G., Valle G.;
RT *Characterization of 16 novel human genes showing high similarity to
RT yeast sequences*.*
RL Yeast 18:69-80(2001).
CC 1- SIMILARITY: BELONGS TO THE ATP-BINDING TRANSPORT PROTEIN FAMILY
CC (ABC TRANSPORTERS). EF-3 SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
109 EMBL; AL050291; CA843392.1; -
109 EMBL; AF261091; AAG13902.1; -
109 EMBL; AF005016; CA06290.1; -
109 InterPro: IPR001617; -
109 Pfam: PF00005; ABC_tran; 2.
109 PROSITE: PS00211; ABC_TRANSPORTER; 1.
109 K1P-Binding; Transport; Repeat.
109 NP_BIND 118 125
109 NP_BIND 430 437
109 NP_BIND 39 39
109 NP_BIND 74 74
109 NP_BIND 623 623
109 NP_BIND 71290 MM; 702A9686RFB061AE CRC64;
109 NP_BIND 623 AA; 71290 MM; 702A9686RFB061AE CRC64;

```

Query Match 3.9%; Score 97.5; DB 1; Length 623;

```

Best Local Similarity 19.4%; Pred. No. 5.3;
Matches 116; Conservative 82; Mismatches 172; Indels 229; Gaps 34;
QY 13 ALPLASMIFFVFNQNSKWSALNISLHWNNSKSLFPPTPL----- 59
Db 72 AVTYGLAS-----HNSDVF-HIINSLTFH-----GQLSDPTKLELSGRRYGLGL 119
QY 60 -----ISLAPLETETELRIKELIE-KLDQIIPRPF-----HVNITTSATHTATIL 105
Db 120 NGIGKSMLSNAIGKREVPPIPEHIDYHLTREMPPSDKPLCHQVEVTERAMLEAEARL 179
QY 106 NPDQYCGDGLHLEVRDHLGRRKQYGDPLRAKSSPALMAGS-----GRVDF 158
Db 180 AHEDAEC-EKMLELIERLEL-----DADKAMKRSRLTGLGFTFAMQRRKLDF 229
QY 159 NNGIYLVSTLFEVGGVSL-----LLTIHSEGV-----SALMSARN-QGYDRI-- 203
Db 230 SCG-----NRMRVALRALTFRPFMLLDEPTNHLDDACVMLEELTKFRILYL 280
QY 204 --FTGQFVNG--TSQVHSEGLILNTNAELCYLDNRQO-----EGF----- 241
Db 281 VSHSDPLNGVCTNIINHH-----NKKLTYTGNYDQYKTRLELEENQKRHMEOD 333
QY 242 -----YCVRPOH-----MPCALTHMSKKNKVS----- 266
Db 334 QIAHKNYIARFQHGSAKLARQAOSKEKTLQKKMASSGLTERVSDKTLSPFPCKIPP 393
QY 267 -----LSQKSLPERSNVGEIMEKFNII-----SVSKNTLK-----S 301
Db 394 PVIVQVNSFKTKNGKPCITNNLEFEGIDLTALVAGNGAGSKTLKLTGELLPTDGM 453
QY 302 VDLAESK-----LQHLAVDDLRNINIQWKYCYPLIGSMYSKEMEYLTGAIDRT 354
Db 454 IRKSHVKGRRYHOHQBOLDLDS--PLEWMAKCP-----EIKKEMKRIIGRY 503
QY 355 G--GEKNTYIV--ISLQGHFR-----PPIDVFIRALNKAHLOHLLRSPTNVI 402
Db 504 GLTGKQGVSPIRNLDSQKRCVCLAMLANQNHMLFLEDPIN-----HLDLETIDAL-- 555
QY 403 IKTENIREKYNDAERPSDFHGYI-----OYLINDIFODLSV--SIIDAW--DITAY 451
Db 556 -----ADAINFEESGMALVSHDFRLIOOVAOEIWCMEKOTITKMGDI-LAY 601
RESULT 12
PINH_ARATH
ID PINH_ARATH STANDARD: PRT; 988 AA.
AC G9XGM1; 049256;
DT 01-OCT-2000 (Rel. 40, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE PINHEAD PROTEIN (ZWILLE PROTEIN).
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; Eurosids II;
OC Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN=CV, LANDSBERG RECTA;
RX MEDLINE=99094873; PubMed=9876176;
RA Lynn K., Fernandez A., Alda M., Sedbrook J., Tasaka M., Masson P.,
RA Barton M.K.;
RT *The PINHEAD/ZWILLE gene acts pleiotropically in Arabidopsis
RT development and has overlapping functions with the ARGONAUTE1 gene.*;
RL Development 126:469-481(1999).
RN 12
RP SEQUENCE FROM N.A.
RC STRAIN=CV, LANDSBERG RECTA;
RX MEDLINE=98169396; PubMed=9501101;
RA Mousian B., Schoof H., Haacker A., Jurgens G., Laux T.;
RT *Role of the ZWILLE gene in the regulation of central shoot meristem
RT cell fate during Arabidopsis embryogenesis.*;

```

```

RL EMBL J. 17.1799-1809(1998).
CC -1- FUNCTION: DEVELOPMENTAL PROTEIN REQUIRED FOR RELIABLE FORMATION OF
CC PRIMARY AND AXILARY SHOOT APICAL MERISTEMS. MAY BE A COMPONENT OF
CC A HYPOTHETICAL MERISTEM FORMING COMPETENCE FACTOR.
CC -1- SIMILARITY: BELONGS TO THE ARKONAUDE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sdb.ch/announce/
CC or send an email to license@sdb-sdb.ch).
CC -----
DR EMBL: AF154272; AAD40098.1; -
DR EMBL: AJ223508; CA11429.1; -
DR Mendel: 24584; Arath:3192;24584.
DR InterPro: IPR003100; -
DR InterPro: IPR003165; -
DR Pfam: PF02170; ZAP: 1.
DR Pfam: PF02171; PLY: 1.
KW Developmental protein.
FT CONFLICT 475 479 PROBE -> AEGOR (IN REF. 2).
FT CONFLICT 671 671 N -> D (IN REF. 2).
SQ SEQUENCE 988 AA; 110867 MW; 32BBB349C613DA20 CRC64;

Query Match 3.9%; Score 97; DB 1; Length 988;
Best Local Similarity 19.9%; Pred. No. 11;
Matches 66; Conservative 45; Mismatches 127; Indels 94; Gaps 15;

QY 212 TSOVHSEGLINTNAELCOYLNRDQEGFYCVRPQIMFCALTHMSKKNKVSYSLSKOE 271
DB 391 TTQPTRELMPPVDENCTMKSVLEYFOEMYGFTIQTHTLPCLOV---GNQKASTLPMEA 446
QY 272 KSLFERNVGEVEIMEKFNITISV-KCNTLSYDLHBSGKLHOLAVLD----- 319
DB 447 CKIVGGRYTKRLNEKOITALKYTCQRPDRNDILKTGYHN-ATQDPYAKEGEMNIS 505
QY 320 -----RNINQOMKY-----CYPLISMTYSYKEN----- 344
DB 506 EKLASVEARILPAPMLKYHENGSKERDCLPQVGOMNMKNKINQMTVSFMACVNFSRSVQ 565
QY 345 EYLTRAIDRTGKKNVTIVISLGOHFRFP-IDVTRRALVNHAKIOHLILRSP----- 397
DB 566 ENVARGCNELGQMEV-----SGMEFNPBPVPIISARPDVEKALKHYHNSMKTKGK 621
QY 398 --DIMVIKTEINREMYNDAREFSDH-GYI-QYIINDIFQ-----DLSYST--- 441
DB 622 EIEILLAILPDNNGSIYGLKRICETELGLISQCCLTTHYKISKQYLANVSLKINXMG 681
QY 442 -----IDA-----MDI-TIAYGTNNVHP 458
DB 682 GRNTVLVDALSCRIDPLVSDIPTIFIGADYTHP 713

RESULT 13
GLCB SYNTP7 STANDARD: PRT: 773 AA.
AC P16934;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 01-FEB-1995 (Rel. 31, Last annotation update)
DE 1,4-ALPHA-GLUCAN BRANCHING ENZYME (EC 2.4.1.18) (GLYCOGEN BRANCHING
DE ENZYME).
GN GLCB.
OS Synechococcus sp. (strain PCC 7942) (Anacystis nidulans R2).
OC Bacteria; Cyanobacteria; Chroococcales; Synechococcus.
RX NCBI_TaxID=1140;
RN [1]
RP SEQUENCE FROM N. A.
RA MEDLINE=90323609; PubMed=2142668;
RA Kiel J.A.K.M., Boels J.M., Belman G., Venema G.;

```

```

RT "Nucleotide sequence of the Synechococcus sp. PCC7942 branching
RT enzyme gene (glbB): expression in Bacillus subtilis."
RL Gene 89:77-84(1990).
CC -1- CATALYTIC ACTIVITY: FORMATION OF 1,6-GLUCOSIDIC LINKAGES OF
CC GLYCOGEN.
CC -1- PATHWAY: THIRD STEP IN GLYCOGEN BIOSYNTHESIS.
CC -1- SUBUNIT: MONOMER.
CC -1- MISCELLANEOUS: THE TEMPERATURE FOR OPTIMAL ACTIVITY IS
CC APPROXIMATELY 35 DEGREES CELSIUS.
CC -1- SIMILARITY: BELONGS TO FAMILY 13 OF GLYCOSYL HYDROLASES, ALSO
CC KNOWN AS THE ALPHA-AMYLASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sdb.ch/announce/
CC or send an email to license@sdb-sdb.ch).
CC -----
DR EMBL: M31544; AAB39038.1; -
DR PIR: J00550; J00550.
DR InterPro: IPR000461; -
DR Pfam: PF00128; Alpha-amylase 1.
KW Glycogen biosynthesis; Transferase; Glycosyltransferase.
FT INIT-MET 0
FT ACT-SITE 439 439 BY SIMILARITY
FT ACT-SITE 492 492 BY SIMILARITY
FT ACT-SITE 560 560 BY SIMILARITY
SQ SEQUENCE 773 AA; 89063 MW; 52BA17CA337BF57 CRC64;

Query Match 3.9%; Score 96.5; DB 1; Length 773;
Best Local Similarity 19.4%; Pred. No. 8.6;
Matches 109; Conservative 77; Mismatches 109; Indels 101; Gaps 10;

QY 33 VW-SALNLSI-SLHW-----NNSKSLFPKTPILSKLPTETELRYE-- 117
DB 151 VMASAHNVSIIIGDPSNWDGKRGHOMARNSGIMELF--IPETIYGAAKYKEIKKNQGHY 204
QY 77 EKLD-----QIIPRPPTHVNTTSAHSTAILLPRPTYKGLQHLLEFQHLGK 341
DB 209 EKSDPYQFOQEVFRP-----TASIVADLDRTYWD-----ADMLEKRR 249
QY 132 Q-----YSGDFLRARMSPALMAGASGK-----VTITNG----- 314
DB 247 HQEPLQPISTVEYVHLGSMHA--SSDAIATDAQCKLPVPVADLKTGAKFLTYREL 304
QY 164 -IYSFTLFWEGGVSLILL--HPSGVSALMSANQGY---DVIYFGQVYVTSVH 219
DB 305 RLIPYVLDL-GYSHIELLPILAHPDQ----SMGYVGYVYAAISKYSPEIFMFCVJCH 400
QY 217 SE-CGLILN-----TNAELCOYLNRDQEGFYCVRPQIMFCALTHMS 269
DB 361 QNGIGVILDWVPGHFPKDGHLAFEDGTHLYEHAOSRGE-----HREMTIVFNV 411
QY 260 KKKVSYLSKQKSLFERSNVGEIMEKFNITISVSKNTLKSVDL-----HESKL 310
DB 413 RHVBNFPLAANLLEFWFKYH-----DGIKVDAYASMLYIDNKKGEYINNEYGR 404
QY 311 QHOLAVLDLRNIN-----IQOMKYCYPLISMTYSYKEN----- 441
DB 465 ENLEADLEIYVNLHLSYPPGALSIAESTSWPVSMPY---VGLGIFLKWNNGM 520
QY 348 TRAIDRTGKKNVTIVISLGOHFRFP-----PIDVTRRALVNHAKIOHLILRSPDTMY 402
DB 521 HDMLD-----YFSMDPWRPQROHQNNNTFSIYWARFSENMALSH-----DEVH 504
QY 403 IKTEINREMYNDA-EFESDPHGYIOYLLI---KDIQDLSVSIIDAMTI 447
DB 565 GKSNIIGKMPDEOMKAFANRLCLGLGMETHGKKTLEPMGERGQMAEVNV 614

```

RESULT 14

GLG_SYN3
ID GLG_SYN3 STANDARD: PRT: 770 AA.
AC P52281:
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE 1,4-ALPHA-GLUCAN BRANCHING ENZYME (EC 2.4.1.18) (GLYCOGEN BRANCHING ENZYME).
GN GLG OR SLL0158.
OS Synechocystis sp. (strain PCC 6803).
OX Bacteria: Cyanobacteria; Chroococcales; Synechocystis.
NCB Taxid:1148.
KN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-96127529; PubMed-8590279;
RA Kaneo T., Tanaka A., Sato S., Kotani H., Sazuka T., Miyajima N., Sugita M., Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis sp. strain PCC6803. 1. Sequence features in the 1 Mb region from map positions 648 to 928 of the genome".
RT DNA Res. 2:157-166(1995).
CC -1- CATALYTIC ACTIVITY: FORMATION OF 1,6-GLUCOSIDIC LINKAGES OF GLYCOGEN.
CC -1- PATHWAY: THIRD STEP IN GLYCOGEN BIOSYNTHESIS.
CC -1- SUBUNIT: MONOMER (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO FAMILY 13 OF GLYCOSYL HYDROLASES, ALSO KNOWN AS THE ALPHA-AMYLASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC -----
DR EMBL: I61999; BAA10073.1; -
DR InterPro: IPR000461; -
DR Pfam: PF00128; alpha-amylase; 1.
KM Glycogen biosynthesis: Transferease; Glycosyltransferase.
FT ACT_SITE 433 433 BY SIMILARITY.
FT ACT_SITE 486 486 BY SIMILARITY.
FT ACT_SITE 554 554 BY SIMILARITY.
SQ SEQUENCE 770 AA: 89527 MW: A435ACAT703FABA CRC64:

Query Match 3.94; Score 96; DB 1; Length 770;
Best local Similarity 18.74; Pred. No. 9.3;
Matches 102; Conservative 84; Mismatches 201; Indels 158; Gaps 29;

```

07 33 VM--SALNLSI--SLHYW-----NNSKSLFRTPLISLKLTEELRIKE--II 76
08 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
10 146 VMAFANRNVSIIGDFNNMNGHLOMKRRNNMWELF--PELGVSTGYEELKRMEEHII 203
07 77 EKLD-----QOIPRPDPTHVNTTSATSHSTATILNPDIYCRGDQHLILEYR--DHLG 128
08 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
10 204 EKTDPYGFYQEVKPK-----TASIVADLDGVOMHDE--DWLEARRTSPLS 247
07 129 R---RKQYGDPLFARMSSPALMACAGS--KYTDNNGTIYSLTFLNF----- 172
08 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
10 248 KPVSYVELHLSGMLHTAYDEPVYTLGSEVPEYSENNMG--ARFLYVELVDKLIPIY 304
07 173 ---GVVSLSLLI--HPSEVSALMSARNNGY-----DRVIFGQ--FVNGTSQV 215
08 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
10 405 KRLATYTHIELPLIAEHFPDG---SMQYGTGYAPTSRFQSPEDFYVDCCHNGIGVI 361
07 216 -----HSEGLILNTVAELCOYLDNRQDEFFYCVRRQNHPCALTHMYSNKK 263
08 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
10 462 IDKVTGHEPKDCH--GLAFDGTHTLIEKGPR-----KEHKEGSLTFYVNGRNEV 410
07 264 VSYLSKQKSLFERSNV--GVETMEKFNITSVSKCNITLKYVDLHESQKLOHLDVADLRNI 322
08 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
10 | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

Db 411 RNELVANALFNEFKYHIDMRVAVASMLYDCREGEEMVANEYGRNLEADPLRQV 470
QY 323 N-----IOMQKCYPIIGSMITSYK-EMELTALADRTGGEKN 359
Db 471 NSVVSYPPIGLIAESTSNWVSWPT-----VGGGGLNKKMNGMHMDLD----- 519
QY 360 TYIVYSLQOHRFP-----PIDVLRALNKAIOHLIASPDTWYIKTENIREMYND 414
Db 520 ---YPSMDPMPROFHONSTFESMWNSSENVMLASH-----DEVVGSNNLKMKGPD 570
QY 415 A-EFSPDHGYOYLII-----KDIPODLSIIDDADITIAVGNNHPPQAVYGNQINI 469
Db 571 EMQKXANRALFTYMTHTGKKTMEKSEKQSENNVWGDLEWHLLNPPHQLKOPFT 630
QY 470 LNTI 474
Db 631 ELNHL 635

RESULT 15

RPOD_AGABT
ID RPOD_AGABT STANDARD: PRT: 1102 AA.
AC P33539:
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE PROBABLE DNA-DIRECTED RNA POLYMERASE (EC 2.7.7.6).
OS Agaricus bisporus.
OG Mitochondrion.
OG Plasmid pEM.
OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Agaricales;
OX Agaricaceae; Agaricus.
NCB Taxid:5343;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 24666 / AG4;
RX MEDLINE-91347410; PubMed-1879001;
RA Robison M.M., Royer J.C., Horgen P.A.;
RT "Homology between mitochondrial DNA of Agaricus bisporus and an internal portion of a linear mitochondrial plasmid of Agaricus bisporus".
RT Curr. Genet. 19:495-502(1991).
CC -1- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS SUBSTRATES.
CC -1- CATALYTIC ACTIVITY: N NUCLEOSIDE TRIPHOSPHATE - N PYROPHOSPHATE + RNA(N)
CC -1- SIMILARITY: BELONGS TO THE PHAGE AND MITOCHONDRIAL RNA POLYMERASES FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X63075; CAA44799.1; -
DR PIR: S28104; S28104.
DR InterPro: IPR002092; -
DR Pfam: PF00940; RNA_pol; 1.
DR PROSITE: PS00489; RNA_POL_PHAGE_2; 1.
DR PROSITE: PS00900; RNA_POL_PHAGE_1; 1.
KM Transferase: Transcription; DNA-directed RNA polymerase;
KM Mitochondrion; Plasmid.
FT ACT_SITE 734 734 BY SIMILARITY.
FT ACT_SITE 804 804 BY SIMILARITY.
FT ACT_SITE 980 980 BY SIMILARITY.
SQ SEQUENCE 1102 AA: 126664 MW: F72FDD308D1AB5B4 CRC64:

Query Match 3.88; Score 95.5; DB 1; Length 1102;

Best Local Similarity 19.4%; Pred. No. 17; Matches 99; Conservative 80; Mismatches 153; Indels 177; Gaps 27;

```
OY 30 STKWSALN-----LSLSLHYNNSTSLPPTPLSLKPLTETELR-IKEI 75
      |:::|
      |:::|
      |:::|
Db 524 SEMLYSSINLYNSIKFKIKKAMLNFIILAEWNNDSKLFKGYNM--LKPILETDSKEIKE- 580
      |:::|
      |:::|
      |:::|
OY 76 IEXLDQIIPPRPTVNTTSATSTATILNPRDYTCRGDQHLILEVRDHLGRKKY-- 133
      |:::|
      |:::|
      |:::|
Db 581 -EKYS-----NSKYLISNIISLASLKDNEFY-----LPYVDFGRVYPLSNYIS 627
      |:::|
      |:::|
      |:::|
OY 134 --GGDFLRARMSPALMAGASGVTDPNNGTYLVSFLLFMEGQVSLSLLIHPSEGSAL 191
      |:::|
      |:::|
      |:::|
Db 628 YOGDGLAR-----SLILFADTKCVLNNSGKEGLN-----V 657
      |:::|
      |:::|
      |:::|
OY 192 WSAKNOCYDRVIPTGQF--VNGTSQVHSEGLILMTNALCOYL-DNRQDEGFYCVRPOH 248
      |:::|
      |:::|
      |:::|
Db 658 YLANLAGYDKLPWSEHLTKVDGIIKEYLESNEISNT-----KIYEDNIDKIS----- 704
      |:::|
      |:::|
      |:::|
OY 249 MPCAALTHTWSKMKKVSYSLSKQESLFEPSNNGVEIMEKFNTISVSKCNTLKSVDLHESG 308
      |:::|
      |:::|
      |:::|
Db 705 EPFOFISIMYAKLLISISN-----PKANISNPIL-----FDASCSCG 739
      |:::|
      |:::|
      |:::|
OY 309 KLOHQLAVDLDR---NINIQW---KYCYP-----LIGSMTYSYKEMEX 346
      |:::|
      |:::|
      |:::|
Db 740 -IQHIALLTLEKELASNVNLTDSNPKEDYPODFTYALEKIRDKLINSIDITELRDITL 798
      |:::|
      |:::|
      |:::|
OY 347 LTRAIIDRTGGEKNVYVI-----SLGQHFRRPPIDVIFIRALNVHKAIOHLILRSPDT 399
      |:::|
      |:::|
      |:::|
Db 799 NRKII-----KRSYMTIPYINISMAIGER-----LMEHFTVTKVLYKRYVYIPGS 843
      |:::|
      |:::|
      |:::|
OY 400 MVIKTEINIREMYNDAERFSDFHGYYOYLIIK-----DIPQDLSVSIID 443
      |:::|
      |:::|
      |:::|
Db 844 ATI-----SSKDYLDYSKYGOLCKIITYVLTKEPLSKLNSNFPESMIDIFYKLNIP1-- 897
      |:::|
      |:::|
      |:::|
OY 444 AW-----DITAGTNNVHPQHVVGNOIN 468
      |:::|
      |:::|
      |:::|
Db 898 TWVTPSGGLIKY-TNKKRPQKVTSVLN 925
      |:::|
      |:::|
      |:::|
```

Search completed: October 12, 2001, 16:05:40
Job time: 183 sec